165

DEPOSITORY
17 OCT 1977

NAVAL RESEARCH LOGISTICS QUARTERLY

JUNE 1977
VOL. 24, NO. 2

OFFICE OF NAVAL RESEARCH

NAVSO P-1278

407-B

NAVAL RESEARCH LOGISTICS QUARTERLY

EDITORS

Murray A. Geisler
Logistics Management Institute

W. H. Marlow
The George Washington University

Bruce J. McDonald
Office of Naval Research

MANAGING EDITOR

Seymour M. Selig
Office of Naval Research
Arlington, Virginia 22217

ASSOCIATE EDITORS

Marvin Denicoff
Office of Naval Research

Alan J. Hoffman
IBM Corporation

Neal D. Glassman
Office of Naval Research

Jack Laderman
Bronx, New York

Thomas L. Saaty
University of Pennsylvania

Henry Solomon
The George Washington University

The Naval Research Logistics Quarterly is devoted to the dissemination of scientific information in logistics and will publish research and expository papers, including those in certain areas of mathematics, statistics, and economics, relevant to the over-all effort to improve the efficiency and effectiveness of logistics operations.

Information for Contributors is indicated on inside back cover.

The Naval Research Logistics Quarterly is published by the Office of Naval Research in the months of March, June, September, and December and can be purchased from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402. Subscription Price: $11.15 a year in the U.S. and Canada, $13.95 elsewhere. Cost of individual issues may be obtained from the Superintendent of Documents.

The views and opinions expressed in this Journal are those of the authors and not necessarily those of the Office of Naval Research.

Issuance of this periodical approved in accordance with Department of the Navy Publications and Printing Regulations, P-35 (Revised 1-74).

A PRIORI ERROR BOUNDS FOR PROCUREMENT COMMODITY AGGREGATION IN LOGISTICS PLANNING MODELS*

A. M. Geoffrion

Graduate School of Management
University of California, Los Angeles
Los Angeles, California

ABSTRACT

A complete logistical planning model of a firm or public system should include activities having to do with the procurement of supplies. Not infrequently, however, procurement aspects are difficult to model because of their relatively complex and evanescent nature. This raises the issue of how to build an overall logistics model in spite of such difficulties. This paper offers some suggestions toward this end which enable the procurement side of a model to be simplified via commodity aggregation in a "controlled" way, that is, in such a manner that the modeler can know and control in advance of solving his model how much loss of accuracy will be incurred for the solutions to the (aggregated) overall model.

I. INTRODUCTION

In this paper the term "procurement" is used in a broad sense that includes materials management of raw materials and parts for a manufacturing firm, the acquisition of goods for subsequent distribution by a wholesale firm, the procurement of supplies and materials by a service organization system, and similar situations. The essential point is that we are addressing the "initial" rather than the "final" stage of a logistics system. See, for instance, the recent book by D. Bowersox [2] which makes the distinction in terms of material management (supplier-oriented) and physical distribution management (customer-oriented).

Whereas it is the large number of customers and their ordering idiosyncrasies that tend to make the final stage of a logistics system hard to model, it is the large number of suppliers and items and sometimes the constantly changing patterns of procurement that frequently make the initial stage difficult to model. Aggregation of customers on a geographic basis into customer zones and aggregation of delivered products (or services) into product groups are commonly used to simplify the final stage of a logistics planning model. Similar aggregations can be used to simplify the initial stage, but satisfactory simplifications may be more difficult to achieve because of the influence of differential supply costs among suppliers and the greater degree of uniqueness as to which suppliers provide what. These influences seem to call for a relatively greater amount

*This research was partially supported by the National Science Foundation and the Office of Naval Research.

of detail to be preserved in the procurement stage of a planning model. Unfortunately, this could require the preparation of unduly detailed procurement forecasts—which suppliers will be able to supply what items at what prices in what annual quantities. The difficulties of assembling this data could be out of proportion to the relative importance of procurement as a component of the total logistics planning model. Even worse, it may not be sensible to impose strict model control in the traditional linear programming sense over procurement activities at so great a level of detail.

A reasonable response to these possible difficulties is to take a more flexible attitude toward the modeling of procurement than is customary among devotees to mathematical programming. Namely, look upon the procurement pattern as an aspect of the problem that is partly given objectively and partly under the analyst's control as though it were a policy parameter. View the procurement pattern as something whose influence is as much to be understood as it is to be "optimized."

The aim of this paper is to provide a rigorous framework within which this flexible modeling attitude can be exercised. We are particularly interested in à priori error bounds concerning the accuracy of the full logistics planning model as it is influenced by aggregating procurement items. So far as we are aware, our results along these lines are without precedent.

A companion paper [5] develops similar results in the context of customer aggregation.

II. MODELING STRATEGIES

As a point of departure, consider the following logistics planning model.

Planning Model P

(1) $$\underset{x,y,z}{\text{minimize}} \sum_{ijk} c_{ijk}x_{ijk}+F(y, z)$$

(2) $$\text{subject to } \underline{S}_{ij} \leq \sum_{k} x_{ijk} \leq \bar{S}_{ij}, \text{ all } i\, j$$

(3) $$\sum_{j} x_{ijk} = \sum_{l} D_{il}y_{kl}, \text{ all } i\, k$$

(4) $$\sum_{k} y_{kl} = 1, \text{ all } l$$

(5) $$x_{ijk} \geq 0, \text{ all } i\, j\, k$$

(6) $$y_{kl} \geq 0, \text{ all } kl \text{ and } (y, z) \in \Omega.$$

The following interpretations will be used:

i=indexes procurement *items* (raw materials, parts, finished goods, etc.).

j=indexes geographical *procurement zones*.

k=indexes the *facilities* being supplied.

l=indexes *customers*.

x_{ijk}=a variable giving the annual amount of item i procured from zone j for facility k.

y_{kl}=a variable giving the fraction of the annual needs of customer l (for goods or services) satisfied by facility k.

z=a vector of additional (possibly logistical) variables.

c_{ijk}=unit cost of procurement plus transportation associated with the flow x_{ijk}.

$F(y, z)$ = the total annual costs associated with (y, z) exclusive of procurement and inbound transportation (typically, facility-related costs plus outbound transportation costs).

$\underline{S}_{ij}$ ($\overline{S}_{ij}$) = a lower (upper) limit on the annual amount of item i procured from zone j (partly given and partly at the analyst's discretion).

D_{il} = the amount of item i required to satisfy the total annual needs of customer l.

Ω = a constraint set that must be satisfied by (y, z).

It is understood that a list L_x of allowable triples (i, j, k) is given to reflect which procurement zones can provide which items to which facilities, and that all summations and constraint enumerations run only over allowable combinations. For instance, the enumeration in (2) over "ij" runs over the pairs $(\bar{i}, \bar{j})$ such that $(\bar{i}, \bar{j}, k) \in L_x$ for some k. Similarly, a list L_y is given which specifies which facilities can serve which customers.

Constraints (2) control the procurement pattern. An historical procurement pattern (or some other preconceived pattern) can be enforced by taking corresponding $\underline{S}_{ij}$ and $\overline{S}_{ij}$'s to be the same or nearly the same. The latitude for departure from the preconceived pattern increases as $\overline{S}_{ij} - \underline{S}_{ij}$ increases. A necessary condition for feasibility is

$$\sum_j \underline{S}_{ij} \leq \sum_l D_{il} \leq \sum_j \overline{S}_{ij} \text{ for all } i. \tag{7}$$

The objective function (1) gives the total cost associated with logistical activities. We have already discussed (2). Constraints (3) specify that each facility must receive exactly enough of each item to satisfy the needs of the customers it serves. This requires that the goods or services demanded by each customer can be converted into corresponding requirements for the constituent items (it is immaterial whether the facilities do manufacturing or distribution or service or some combination thereof). Constraints (4) specify that the full needs of each customer must be satisfied. Constraints (5) and (6) impose whatever other requirements on the variables may be needed for system feasibility.

Observe that for fixed y and z, the optimization over x separates into independent subproblems for each i—each a slight generalization of the classical minimum cost transportation problem.

Because of the complete generality of F and Ω, the model could be set up to determine the least cost facility locations satisfying a desired level of customer service. Normally this would require that F be discontinuous in order to accommodate fixed costs, or some binary z-variables could be introduced to achieve the same effect. The model could also be set up to provide for multiple commodities flowing to customers from the facilities, unique assignment of customers to facilities for certain commodities, and many other problem features. We prefer to leave the model in its general form (1)–(6) because these and many other special cases are thereby treated simultaneously with minimum notational complexity.

The model as stated is actually just a point of departure for the models we actually wish to study. Its chief shortcoming is that it may involve too great a level of detail regarding procurement from the viewpoint of policy and also sheer size. Consider first the policy aspect. Model P places limits on the procurement pattern (via (2)) on an item-by-item basis. Except for items of major importance, this seems like an excessive degree of control and may not even be meaningful in situations where suppliers are changed frequently on the basis of current price and availability. It

would make more sense when there are many items of small importance to aggregate some of the constraints in (2). Suppose this is done for some subset I of items. The result is

Planning Model P_I

the same as planning model P, except that (2) is replaced by

$$\underline{S}_{ij} \leq \sum_k x_{ijk} \leq \overline{S}_{ij}, \text{ all } ij \text{ with } i \notin I \tag{2.1}$$

$$\underline{S}_{I,j} \leq \sum_{\substack{i \in I \\ k}} x_{ijk} \leq \overline{S}_{I,j}, \text{ all } j, \tag{2.2}$$

where

$$\underline{S}_{I,j} \triangleq \sum_{i \in I} \underline{S}_{ij} \text{ and } \overline{S}_{I,j} \triangleq \sum_{i \in I} \overline{S}_{ij}. \tag{8}$$

This version seems more reasonable from a policy standpoint in that the procurement pattern for items I is now stipulated on an aggregate basis. The numbers $\underline{S}_{I,j}$ and $\overline{S}_{I,j}$ would be interpreted rather freely since their formal constituents $\underline{S}_{ij}$ and $\overline{S}_{ij}$ might be poorly known or perhaps even ill-defined.

There is, of course, a natural generalization of P_I that aggregates the procurement pattern constraints for *several* subsets of items. The analysis of this generalization is a simple extension of the results to be obtained for P_I (see the Remark in Appendix 1).

Model P_I is better from a policy standpoint but it still may be too large. The number of variables is unchanged, although the number of type (2) constraints has diminished. Moreover, a possible new difficulty arises in that the mathematical structure of P_I is more complex than that of P. This is due to the fact that aggregating the type (2) constraints over $i \in I$ has the effect of coupling together what previously was a collection of independent transportation-like subproblems in the x-variables when y and z are fixed. The new coupling tends to diminish the computational effectiveness of solution methods that exploit the natural separation into subproblems when y and z are held fixed temporarily (e.g., methods based on Benders decomposition [4]). The nice structure of P could be restored, and the size of P_I much reduced, by completing the aggregation with respect to items I begun in the passage from P to P_I. This involves replacement of the variables x_{ijk} with $i \in I$ by aggregate variables ξ_{jk}, say, so that the following single transportation-like subproblem replaces the coupled subproblems of P_I for fixed y:

$$\underset{\xi \geq 0}{\text{Minimize}} \sum_{jk} b_{jk} \xi_{jk}$$

subject to

$$\underline{S}_{I,j} \leq \sum_k \xi_{jk} \leq \overline{S}_{I,j}, \text{ all } j \tag{2.2A}$$

$$\sum_j \xi_{jk} = \sum_{\substack{i \in I \\ l}} D_{il} y_{kl}, \text{ all } k, \tag{3.1}$$

where the b_{jk}'s are plausible surrogates for the c_{ijk}'s over $i \in I$. Variable ξ_{jk} is interpreted as a surrogate for $\sum_{i \in I} x_{ijk}$, and (3.1) is interpreted as requiring facility k to receive enough of the items in I to meet its needs in the aggregate.

This further aggregation of P_I leads to

Planning Model $\tilde{P}_{I,b}$

$$\text{(9)} \qquad \underset{x,y,z,\xi}{\text{Minimize}} \sum_{i \notin I} \sum_{jk} c_{ijk}x_{ijk} + \sum_{jk} b_{jk}\xi_{jk} + F(y, z) + L(y; b)$$

subject to

$$\text{(2.1)} \qquad \underline{S}_{ij} \leq \sum_{k} x_{ijk} \leq \overline{S}_{ij}, \text{ all } ij \text{ with } i \notin I$$

$$\text{(2.2A)} \qquad \underline{S}_{I,j} \leq \sum_{k} \xi_{jk} \leq \overline{S}_{I,j}, \text{ all } j$$

$$\text{(3.1)} \qquad \sum_{j} \xi_{jk} = \sum_{l} D_{I,l} y_{kl}, \text{ all } k$$

$$\text{(3.2)} \qquad \sum_{j} x_{ijk} = \sum_{l} D_{il} y_{kl}, \text{ all } ik \text{ with } i \notin I$$

$$\text{(4)} \qquad \sum_{k} y_{kl} = 1, \text{ all } l$$

$$\text{(5.1)} \qquad x_{ijk} \geq 0, \text{ all } ijk \text{ with } i \notin I$$

$$\text{(5.2)} \qquad \xi_{jk} \geq 0, \text{ all } jk \text{ such that } ijk \text{ exists for } i \in I$$

$$\text{(6)} \qquad y_{kl} \geq 0, \text{ all } kl \text{ and } (y, z) \in \Omega,$$

where we define

$$\text{(10)} \qquad D_{I,l} \triangleq \sum_{i \in I} D_{il}$$

and where $L(y; b)$ is some linear function of y designed to "compensate" for aggregation error in spite of the arbitrary choice of b.

Notice that the mathematical structure of $\tilde{P}_{I,b}$ is identical to that of P (with the addition to the objective function of a new term linear in y, which seems innocuous enough). $\tilde{P}_{I,b}$ is smaller in that the items of I have been aggregated together throughout.

The major task at this point is to understand the relationship between P_I and $\tilde{P}_{I,b}$. Our main results in this direction are summarized in the next section.

III. THE RELATIONSHIP BETWEEN PLANNING MODELS P_I AND $\tilde{P}_{I,b}$

As it turns out, a natural choice for the L function exists for which a nearly ideal relationship can be established between P_I and $\tilde{P}_{I,b}$. In particular, an *à priori* bound can be obtained on the difference between their optimal values. Such a bound can be obtained for any choice of b, and in fact furnishes a useful criterion for making this choice.

It will be convenient to refer to the so-called Range function, which is defined for any collection $\{\alpha_1, \ldots, \alpha_n\}$ of scalars as

$$\underset{1 \leq j \leq n}{\text{Range}} \{\alpha_j\} \triangleq \underset{1 \leq j \leq n}{\text{Max}} \{\alpha_j\} - \underset{1 \leq j \leq n}{\text{Min}} \{\alpha_j\}.$$

The notation $v(\cdot)$ will refer to the optimal value of an optimization problem.

MAIN THEOREM: Assume that the same jk links exist for every item in some subset I. Let b_{jk} be chosen arbitrarily for these links, and take the compensation function L to be

$$L(y;\, b)=\sum_{kl}\left(\sum_{i\epsilon I} D_{il}\operatorname*{Min}_{j}\{c_{ijk}-b_{jk}\}\right)y_{kl}. \tag{11}$$

Then

$$v(\tilde{P}_{I,b})\leq v(P_I)\leq v(\tilde{P}_{I,b})+\epsilon_b, \tag{12}$$

where

$$\epsilon_b\triangleq\sum_{l}\operatorname*{Max}_{k}\{\sum_{i\epsilon I} D_{il}\operatorname*{Range}_{j}\{c_{ijk}-b_{jk}\}\}. \tag{13}$$

Moreover, a complete ϵ_b-optimal solution of P_I can be obtained from any optimal solution $(\tilde{x}, \tilde{y}, \tilde{z}, \tilde{\xi})$ to $\tilde{P}_{I,b}$ by using $(\tilde{x}, \tilde{y}, \tilde{z})$ as is and supplementing it by values for the missing x_{ijk} for $i \epsilon I$ according to the disaggregation formula: for all ijk with $i \epsilon I$, put

$$\tilde{x}_{ijk}=\begin{cases}\dfrac{\sum_{l} D_{il}\tilde{y}_{kl}}{\sum_{l} D_{I,l}\tilde{y}_{kl}}\,\tilde{\xi}_{jk} & \text{if } \tilde{\xi}_{jk}>0\\[2ex] 0 & \text{if } \tilde{\xi}_{jk}=0\end{cases} \tag{14}$$

The proof is given in Appendix 1, along with a generalization to the case where several subsets of items are aggregated simultaneously. Extensions which accommodate suboptimal solutions to $\tilde{P}_{I,b}$ are easy to obtain.

This theorem is a satisfying one in a number of respects. First, it allows for an arbitrary aggregation set I subject to the requirement that the items involved have a common set of transportation links (otherwise feasibility difficulties could be encountered in trying to recover a feasible solution to P_I from one of $\tilde{P}_{I,b}$). Second, it allows an arbitrary choice of b, which accommodates any heuristic rule that may be appealing in a particular situation (e.g., some weighted mean of c_{ijk} over $i \epsilon I$). Third, it selects L in such a manner that the aggregated problem is a *relaxation* of the original one in a suitable sense, thereby producing an underestimate of the optimal value of the original problem. Fourth, this underestimate has an error that is known *à priori* to be no larger than a calculable number ϵ_b. Fifth, solving the aggregated problem is guaranteed to furnish a complete ϵ_b-optimal solution to P_I (one can very likely conclude that this solution is ϵ-optimal in P_I for some ϵ smaller than ϵ_b—just take the difference between the objective function (1) evaluated at the feasible solution and the lower bound $v(\tilde{P}_{I,b})$). And sixth, the explicit formula for ϵ_b has a number of valuable applications. We now expand on this last point.

An important question is how one should select b when a compelling heuristic choice is not available. The formula for ϵ_b furnishes a natural criterion: select b to make ϵ_b as small as possible. Happily, this can be converted to a linear programming problem by using standard tricks (mainly the representation of the maximum of a set of numbers as their least upper bound). Thus, *the optimal b can always be calculated by linear programming.*

The ϵ_b-minimizing choice of b can sometimes be obtained analytically if additional assumptions are imposed. For instance, if the D_{il}'s are proportionally the same for i in I at every customer—i.e., if there exist proportions

$$p_i, \text{ where } p_i \geq 0 \text{ for } i \epsilon I \text{ and } \sum_{i \epsilon I} p_i = 1$$

such that

$$\frac{D_{il}}{\sum_{i \epsilon I} D_{il}} = p_i \text{ for all } il \text{ with } i \epsilon I \tag{15}$$

—and $p_{i_0} \geq \sum_{\substack{i \epsilon I \\ i \neq i_0}} p_i$ for some $i_0 \epsilon I$, then it can be shown that the optimal choice of b is to take $b_{jk} = c_{i_0 jk}$ for all jk.

It is of interest to characterize the situations where $\epsilon_b = 0$ is possible. It is shown in Appendix 2 that a necessary and sufficient condition for ϵ_b to equal 0 for some choice of b is that there exist numbers β_{jk} and γ_{ik} such that

$$c_{ijk} = \beta_{jk} + \gamma_{ik} \quad \text{for all } ijk \text{ with } i \epsilon I \text{ and } k \text{ such that it is connected to some } l \text{ with } D_{il} > 0. \tag{16}$$

If this condition holds, then $\epsilon_b = 0$ is achieved by taking $b_{jk} = \beta_{jk}$ for all jk (plus any constant depending only on k) with k such that it is connected to some l for which

$$\sum_{i \epsilon I} D_{il} > 0.$$

The choice of b_{jk} is arbitrary for any k's left over.

When might (16) hold? An important case occurs when item i has a procurement cost γ_i \$/unit, and all items in I have the same unit inbound transportation rate when measured on a per mile basis, say t_I \$/unit-mile. If the distance from j to k is d_{jk}, then

$$c_{ijk} = t_I d_{jk} + \gamma_i \quad \text{for all } ijk \text{ with } i \epsilon I \tag{17}$$

and (16) clearly holds. This case admits an easy generalization that still leaves $\epsilon_b = 0$: t_I can depend on j or k or both, and γ_i can depend on k.

IV. CONCLUSION

We have achieved our goal of providing rigorous guidance to the modeler who wishes to consider aggregating a subset I of items in the procurement portion of a logistics planning model. Assuming that the aggregate constraints (2.2) offer adequate control of the procurement pattern, the modeler can obtain an *à priori* bound from (13) on the amount of suboptimality that will be caused in the model by subsequently collapsing the inbound flows for i in I down to a single transportation-like problem that uses any plausible costs b_{jk} for the aggregated items. It bears emphasis that this bound can be calculated *before* optimizing the aggregated planning model, perhaps using rough preliminary data, and hence can be a useful tool for model design.

The results attained can be used not only to study the effects of aggregation with a predetermined subset I of items, but also to select I itself on the basis of small anticipated aggregation error. This can be done by cluster analysis aimed at finding item subsets for which (16) holds approximately. One way to proceed is based on the following observation. Notice that if (16) holds exactly, then summing over j yields

$$\sum_j c_{ijk} = \sum_j \beta_{jk} + ||j||_i \gamma_{ik} \text{ for } ik,$$

where $||j||_i$ is the number of procurement zones supplying item i. Thus, γ_{ik} can be eliminated in (16) using

$$\gamma_{ik} = \frac{\sum_j c_{ijk}}{||j||_i} - \frac{\sum_j \beta_{jk}}{||j||_i}$$

to obtain

$$(16)' \qquad c_{ijk} - \frac{\sum_j c_{ijk}}{||j||_i} = \beta_{jk} - \frac{\sum_j \beta_{jk}}{||j||_i} \qquad \begin{array}{l}\text{for all } ijk \text{ with } i \epsilon I \text{ and } k \\ \text{such that it is connected} \\ \text{to some } l \text{ with } D_{il} > 0.\end{array}$$

Conversely, (16)′ implies that (16) holds. Hence (16) and (16)′ are equivalent conditions. The obvious clustering approach would be to identify with each item i a linearized vector V^i with typical entry

$$\begin{cases} c_{ijk} - \dfrac{\sum_j c_{ijk}}{||j||_i} & \text{if link } ijk \text{ exists} \\ \text{a large number } M & \text{otherwise.} \end{cases}$$

The V^i-vectors would then be clustered by some standard technique [1] to discover subsets of i for which the V^i's are nearly identical. These subsets of i would identify items which, if aggregated, would tend to have small aggregation error when an approrpiate choice for b is used. In fact, an appropriate choice for b would be a virtual by-product of most standard clustering schemes

A refinement would be to weight the V^i's or its components according to demand or some measure of the likelihood that a given link would actually be selected by the model for use.

REFERENCES

[1] Anderberg, *Cluster Analysis for Applications* (Academic Press, 1973).

[2] Bowersox, D. J., *Logistical Management* (Macmillan, 1974).

[3] Geoffrion, A. M., "Elements of Large-Scale Mathematical Programming," Management Science, *16* (11) 652–691 (July 1970).

[4] Geoffrion, A. M. and G. W. Graves, "Multicommodity Distribution System Design by Benders Decomposition," Management Science, *20* (5) 822–844 (January 1974).

[5] Geoffrion, A. M., "Customer Aggregation in Distribution Modeling," Working Paper No. 259, Western Management Sciences Institute, UCLA (October 1976).

[6] Geoffrion, A. M., "Objective Function Approximations in Mathematical Programming," Mathematical Programming, to appear.

APPENDIX 1: PROOF OF THE MAIN THEOREM

Let $v(\cdot)$ denote the infimal value of any minimizing optimization problem.

Lemma 1 [6]. Consider the two optimization problems

$$(Q) \text{ Minimize } f(w) \text{ subject to } w \epsilon W$$

$$(\tilde{Q}) \text{ Minimize } \tilde{f}(w) \text{ subject to } w \epsilon W,$$

where f and $\tilde{f}$ are real-valued functions bounded below on a non-empty set W. (Interpret (Q) as the "true" problem and $(\tilde{Q})$ as the "approximating" problem in the sense that an approximate objective function $\tilde{f}$ replaces f.) Let $\underline{\epsilon}$ and $\bar{\epsilon}$ be scalars (not necessarily nonnegative) satisfying

$$\text{(A1)} \qquad -\underline{\epsilon} \leq \tilde{f}(w) - f(w) \leq \bar{\epsilon} \text{ for all } w \epsilon W.$$

Then

$$\text{(A2)} \qquad -\underline{\epsilon} \leq v(\tilde{Q}) - v(Q) \leq \bar{\epsilon}$$

and any optimal solution $\tilde{w}$ of $(\tilde{Q})$ is necessarily $(\underline{\epsilon}+\bar{\epsilon})$-optimal in (Q).

Lemma 1 will be applied not to P_I in the role of (Q), but rather to an equivalent version of P_I, namely its "projection" [3] onto the variables y, z, and x with $i \notin I$:

$$(P_I)^* \qquad \underset{x,y,z}{\text{Minimize}} \; F(y, z) + \sum_{\substack{i \notin I \\ jk}} c_{ijk} x_{ijk} + \varphi_I(y)$$

$$\text{subject to (2.1), (3.2), (4), (5.1), (6)}$$

where we define

$$\text{(A3)} \qquad \varphi_I(y) \triangleq \text{Infimum} \quad \sum_{\substack{ijk \\ i \epsilon I}} c_{ijk} x_{ijk} \text{ subject to (2.2) and}$$

$$\sum_j x_{ijk} = \sum_l D_{il} y_{kl}, \text{ all } ik \text{ with } i \epsilon I$$

$$x_{ijk} \geq 0, \text{ all } ijk \text{ with } i \epsilon I.$$

Make the identifications

$$w = \text{the variables of } (P_I)^*$$

$$W = \text{the constraints of } (P_I)^*$$

$$f(w) = \text{the objective function of } (P_I)^*$$

$$\tilde{f}(w) = \text{the objective function of } (P_I) \text{ with } \varphi_I \text{ replaced by } \tilde{\varphi}_I,$$

where $\tilde{\varphi}_I(y)$ is defined as

$$\text{(A4)} \qquad \tilde{\varphi}_I(y) \triangleq L(y;\, b) + \underset{\xi \geq 0}{\text{Inf.}} \sum_{jk} b_{jk} \xi_{jk} \text{ subject to (2.2A) and (3.1)}$$

with L as defined in (11) for arbitrary fixed b. The justification for (A4) is provided by

LEMMA 2: Assume that the same jk links exist for every item in the subset I. Then (A5)

$$\tilde{\varphi}_I(y) \leq \varphi_I(y) \leq \tilde{\varphi}_I(y) + \epsilon_b, \text{ all } (y, z) \text{ satisfying (4) and (6), where } \epsilon_b \text{ is defined as in (13).}$$

Once Lemma 2 is established, conclusion (12) of the Main Theorem is at hand upon applying Lemma 1 using the identifications given above and the obvious facts $v(Q)=v(P_I)^*=v(P_I)$ and $v(\tilde{Q})=v(\tilde{P}_{I,b})$.

PROOF OF LEMMA 2: Introduce a supplementary nonnegative variable ξ_{jk} into (A3) for each jk link in existence for $i \epsilon I$, along with the supplementary constraints

$$\xi_{jk} = \sum_{i \epsilon I} x_{ijk}$$

and the supplementary terms $b_{jk}\,\xi_{jk} - b_{jk}\xi_{jk}$ in the objective function. From (2.2) we see that additional redundant constraints (2.2A) may be added, and from the demand constraints of (A3) we see that (3.1) may be added. Clearly none of this alters the infimal value of (A3). Upon "projection" of the augmented problem onto the ξ-variables, one obtains

(A3)* $$\varphi_I(y) = \underset{\xi \geqq 0}{\text{Infimum}} \sum_{jk} b_{jk}\xi_{jk} + R(\xi, y) \text{ subject to (2.2A), (3.1)}$$

where the remainder term is defined as

$$R(\xi, y) \triangleq \text{Infimum} \sum_{\substack{ijk \\ i \epsilon I}} (c_{ijk} - b_{jk}) x_{ijk}$$

subject to

$$\sum_j x_{ijk} = \sum_l D_{il} y_{kl}, \text{ all } ik \text{ with } i \epsilon I$$

$$\sum_{i \epsilon I} x_{ijk} = \xi_{jk}, \text{ all } jk$$

$$x_{ijk} \geq 0, \text{ all } ijk \text{ with } i \epsilon I.$$

It is easy to verify that

$$\underline{R}(y) \leq R(\xi, y) \leq \overline{R}(y) \qquad \text{for all } (y, z) \text{ satisfying (4) and (6) and } \xi \text{ satisfying (2.2A) and (3.1),}$$

where

$$\underline{R}(y) \triangleq \sum_{kl} \Big(\sum_{i \epsilon I} D_{il} \operatorname*{Min}_j \{c_{ijk} - b_{jk}\}\Big) y_{kl} \equiv L(y; b) \qquad \text{as defined in (11)}$$

$$\overline{R}(y) \triangleq \sum_{kl} \Big(\sum_{i \epsilon I} D_{il} \operatorname*{Max}_j \{c_{ijk} - b_{jk}\}\Big) y_{kl}.$$

Since $\overline{R}(y) - \underline{R}(y)$ clearly is no larger than

$$\sum_l \operatorname*{Max}_k \Big\{\sum_{i \epsilon I} D_{il} [\operatorname*{Max}_j \{c_{ijk} - b_{jk}\} - \operatorname*{Min}_j \{c_{ijk} - b_{jk}\}]\Big\}$$

$$= \sum_l \operatorname*{Max}_k \Big\{\sum_{i \epsilon I} D_{il} \operatorname*{Range}_j \{c_{ijk} - b_{jk}\}\Big\} \equiv \epsilon_b \text{ as defined in (13)}$$

for any $y \geqq 0$ satisfying (4), we have

(A6) $$L(y; b) \leq R(\xi, y) \leq L(y; b) + \epsilon_b \qquad \text{for all } (y, z) \text{ satisfying (4) and (6) and } \xi \text{ satisfying (2.2A) and (3.1).}$$

The desired conclusion (A5) now follows easily from (A3)* and (A6). This completes the proof of Lemma 2.

Finally we come to the second conclusion of the Main Theorem. Let $(\tilde{x}^-, \tilde{y}, \tilde{z}, \tilde{\xi})$ be any optimal solution to $\tilde{P}_{I,b}$ and generate $\tilde{x}^+_{ijk}$ for $i\epsilon I$ according to

$$\tilde{x}^+_{ijk}=\frac{\sum_l D_{il}\tilde{y}_{kl}}{\sum_l D_{I.l}\tilde{y}_{kl}}\tilde{\xi}_{jk}, \text{ all } ijk \text{ with } i\epsilon I.$$

This "any feasible disaggregation of $\tilde{\xi}$" construction is possible because of the assumption that the same jk links exist for all $i\epsilon I$. We must show that $(\tilde{x}^-, \tilde{x}^+, \tilde{y}, \tilde{z})$ is feasible and ϵ_b-optimal in P_I. The verification of feasibility is straightforward. To verify ϵ_b-optimality we need to show

$$\sum_{\substack{ijk\\ i\notin I}} c_{ijk}\tilde{x}^-_{ijk}+\sum_{\substack{ijk\\ i\epsilon I}} c_{ijk}\tilde{x}^+_{ijk}+F(\tilde{y},\tilde{z})\leq v(P_I)+\epsilon_b.$$

This is an obvious consequence of (12) and

$$v(\tilde{P}_{I,b})\leq\sum_{ijk} c_{ijk}\tilde{x}_{ijk}+F(\tilde{y},\tilde{z})\leq v(\tilde{P}_{I,b})+\epsilon_b.$$

This last result, in turn, is a simple consequence of these two facts:

$$\sum_{\substack{ijk\\ i\notin I}} c_{ijk}\tilde{x}^-_{ijk}+\sum_{jk} b_{jk}\tilde{\xi}_{jk}+F(\tilde{y},\tilde{z})+L(\tilde{y};b)=v(\tilde{P}_{I,b}),$$

which holds by the definition of $(\tilde{x}^-, \tilde{y}, \tilde{z}, \tilde{\xi})$, and

$$L(\tilde{y};b)\leq\sum_{\substack{ijk\\ i\epsilon I}}(c_{ijk}-b_{jk})\tilde{x}^+_{ijk}\leq L(\tilde{y};b)+\epsilon_b,$$

which can be simplified to

$$L(\tilde{y};b)\leq\sum_{\substack{ijk\\ i\epsilon I}} c_{ijk}\tilde{x}^+_{ijk}-\sum_{jk} b_{jk}\tilde{\xi}_{jk}\leq L(\tilde{y};b)+\epsilon_b.$$

This completes the proof of the Main Theorem.

REMARK: It is a straightforward matter to generalize the Main Theorem to cover the case where several disjoint subsets of items are to be aggregated, say $I^1, \ldots, I^H$. The analogs of P_I and $\tilde{P}_{I,b}$ should be obvious. Assume for $h=1, \ldots, H$ that the same jk links exist for every item in subset I^h and choose b^h_{jk} arbitrarily for these links. Define

$$L^h(y;b^h)\triangleq\sum_{kl}\left(\sum_{i\epsilon I^h} D_{il}\operatorname*{Min}_j\{c_{ijk}-b^h_{jk}\}\right)y_{kl}.$$

Then

$$v\text{ (analog of }\tilde{P}_{I,b})\leq v\text{ (analog of }P_I)\leq v\text{ (analog of }\tilde{P}_{I,b})+\epsilon_b^H,$$

where

$$\epsilon_b{}^H\triangleq\sum_l\operatorname*{Max}_k\left\{\sum_{h=1}^{H}\sum_{i\epsilon I^h} D_{il}\operatorname*{Range}_j\{c_{ijk}-b^h_{jk}\}\right\},$$

and an ϵ_b^H-optimal solution of the analog of P_I can be constructed in the obvious way. Note that ϵ_b^H is smaller than the tolerance that would be obtained from H successive applications of the original version of the Main Theorem.

APPENDIX 2: NECESSARY AND SUFFICIENT CONDITIONS FOR ZERO AGGREGATION ERROR

PROPOSITION: $\epsilon_b=0$ in expression (13) if and only if there exist numbers γ_{ik} such that

$$c_{ijk}=b_{jk}+\gamma_{ik} \qquad \text{for all } ijk \text{ with } i \in I \text{ and } k \text{ such that it is connected to some } l \text{ with } D_{il}>0.$$

PROOF: It is easy to see that $\epsilon_b=0$ if and only if

$$D_{il} \operatorname*{Range}_j \{c_{ijk}-b_{jk}\}=0 \qquad \text{for all possible } ikl \text{ with } i \in I \text{ (for } ikl \text{ to be possible, } k \text{ must be connected to } l \text{ and } ijk \text{ must exist for some } j)$$

which, by the nonnegativity of D_{il} and of the range function, holds if and only if

(A7) $$\operatorname*{Range}_j \{c_{ijk}-b_{jk}\}=0 \qquad \text{for all possible } ik \text{ with } i \in I \text{ and } k \text{ such that it is connected to some } l \text{ with } D_{il}>0.$$

Now the range function has the property that it vanishes if and only if all of its arguments are identical, and so (A7) holds if and only if numbers γ_{ik} exist such that

$$c_{ijk}-b_{jk}=\gamma_{ik} \qquad \text{for all } ijk \text{ with } i \in I \text{ and } k \text{ such that it is connected to some } l \text{ with } D_{il}>0.$$

A NODE COVERING ALGORITHM*

Egon Balas and Haakon Samuelsson†

Carnegie-Mellon University
Pittsburgh, Pennsylvania

ABSTRACT

This paper describes a node covering algorithm, i.e., a procedure for finding a smallest set of nodes covering all edges of an arbitrary graph. The algorithm is based on the concept of a dual node-clique set, which allows us to identify partial covers associated with integer dual feasible solutions to the linear programming equivalent of the node covering problem. An initial partial cover with the above property is first found by a labeling procedure. Another labeling procedure then successively modifies the dual node-clique set, so that more and more edges are covered, i.e., the (primal) infeasibility of the solution is gradually reduced, while integrality and dual feasibility are preserved. When this cannot be continued, the problem is partitioned and the procedure applied to the resulting subproblems. While the steps of the algorithm correspond to sequences of dual simplex pivots, these are carried out implicitly, by labeling. The procedure is illustrated by examples, and some early computational experience is reported. We conclude with a discussion of potential improvements and extensions.

1. INTRODUCTION

The problem for which this paper presents an algorithm can be stated as follows. Given an undirected graph $G=(V, E)$, where V is the set of nodes and E the set of edges of G, i.e., $E \subseteq V \times V$, find a subset $V^* \subseteq V$ of minimum cardinality, such that all members of E are incident with at least one member of V^*. This problem, called in the literature the node covering problem, can be stated in integer programming format as

$$Z_{NC} = \text{Min } e_p y$$

$$\text{(NC)} \qquad \text{s.t. } A^T y \geq e_q$$

$$y \in \{0,1\}^p$$

where $e_p(e_q)$ is a p-vector (q-vector) of ones, A is the node-edge incidence matrix of the graph $G=(V,E)$, $p=|V|$, $q=|E|$, and T denotes transpose.

---

*This research was supported by the Office of Naval Research and the National Science Foundation. An earlier version of the paper was circulated under [1].

†Research for this article was performed prior to the death of Professor Samuelson in May, 1975.

By the transformation $y'=e_p-y$ we obtain the problem:

$$\text{(NP)} \qquad \begin{aligned} Z_{NP} &= \text{Max } e_p y' \\ &\text{s.t. } A^T y' \leq e_q \\ &y' \in \{0,1\}^p \end{aligned}$$

which is the node packing problem, i.e., the problem of finding a maximum-cardinality independent set of nodes in the same graph G. Thus, if V^* is a minimum node cover, then $V-V^*$ is a maximum independent set of nodes, and $Z_{NC}+Z_{NP}=p$ (see [8]).

If G is the complement of G, i.e., $G'=(V, E')$, where $(i, j) \in E'$ if and only if $(i, j) \notin E$, then (NP) amounts to the problem of finding a maximum-cardinality clique in G'. (A clique is defined to be a complete subgraph.)

The node covering and node packing problems have several important practical applications (see for instance [5, 13, 14, 2]). The most famous one is formulated in terms of the node packing problem and concerns k-related subsets of a set.

Let $S=\{1, \ldots, N\}$ be the set. We then search for a largest number of subsets $S_i \subset S$ of a fixed size $|S_i|=n$, $\forall i$, such that no two sets S_i, S_j have more than k elements in common. This problem is solved by associating nodes of a graph with all subsets of size n, and edges of the graph with all pairs of nodes corresponding to subsets with more than k common elements. If S is a set of treatments in a statistical experiment and S_i are blocks, we have a classical problem in experimental design.

Another interpretation is that S is a set of prospective committee-members and the subsets a number of committees where for some reason no group of k members should serve on more than one committee.

A common problem in information retrieval can be modeled in terms of a maximum clique problem. If pieces of information are represented by nodes of a graph and relations between the former as edges of the same graph, the problem of finding a maximum totally related set of data elements clearly amounts to finding a maximum clique in the graph, or equivalently a maximum node packing in its complement.

Shannon [13] gives the following application to information theory. A set of signals contains a number of pairs that can be confused by a receiver. The problem of finding a largest set of signals to use so as to exclude the possibility of confusion is a node packing problem in a graph, with nodes representing signals, and with edges between those nodes corresponding to signals that can be confused.

A number of classical combinatorial problems such as Gauss' chess problem (place eight queens on a chess-board out of reach of one another) can also be modeled as node covering problems.

Our theoretical interest in the node covering problem, however, derives primarily from the fact that it is one of the few integer programs for which anything at all is known about the structure of the convex hull of integer points feasible to the linear program

$$\text{(LNC)} \qquad \begin{aligned} Z_{LNC} &= \text{Min } e_p y \\ &\text{s.t. } A^T y \geq e_q \\ &y \geq 0 \end{aligned}$$

associated with (NC); and one is of course tempted to try to put this knowledge to use in some solution procedure more efficient than those not using it.

For a close relative of our problem, namely the edge matching problem, which can be stated as

$$Z_{EM}=\text{Max } e_q x$$

(EM) $$\text{s.t. } Ax \leq e_p$$

$$x \in \{0,1\}^p,$$

the convex hull of integer points feasible to the associated linear program (LEM), in which $x \in \{0,1\}^p$ is replaced by $x \geq 0$, has been fully characterized by Edmonds [7], who has also given a polynomially bounded algorithm for solving this problem.

The obvious connection between the edge matching and the node covering problem consists in the fact that the linear programs associated with the two problems, (LEM) and (LNC), are dual to each other, and thus an optimal solution (and, for that matter, any feasible solution) to (LEM) gives a valid lower bound on Z_{LNC}, and hence on Z_{NC}. We will make ample use of this connection and, more importantly, of a less obvious one which will be discussed in the next section; namely, that with any feasible matching one can associate a partial cover, i.e., a solution satisfying a subset of the constraints of (NC), which possesses certain desirable properties.

Unlike for the edge matching problem, for the node covering problem the convex hull H_I of the integer points feasible to (LNC) has so far been only partially characterized. Several classes of facets have been identified by Padberg [12], Chvátal [6], Nemhauser and Trotter [11], Trotter [15], Balas and Zemel [3, 4]. (For a recent survey of this whole area, see Balas and Padberg [2]). The simplest type of nontrivial facets, and the only ones that we will make use of, are the facets associated with maximal cliques. To be more specific, let $Q \subset V$ be any subset of pairwise adjacent nodes of G; then the inequality

$$\sum_{j \in Q} y(j) \geq |Q|-1$$

where $y(j)$ is the variable of (NC) associated with node j, defines a face of H_I, and if Q is of maximum cardinality, then it defines a facet, i.e., a (d-l)-dimensional face of H_I, where $d=\dim H_I$.

Other classes of facets could be used in a way similar to the procedure discussed in this paper. The algorithm to be described below, however, only uses the facets associated with cliques.

A couple of other results, not used by our algorithm, but related to our problem, are as follows. A graph is *chordal* if it has no cycles of size greater than three without chords; and it is a *circle graph* if its vertices can be identified with chords in a circle and its edges with pairs of chords that cross each other. For both cases, polynomially bounded algorithms for the node covering problem have been found by Gavril [9, 10].

In the next section we introduce the concept of a dual node-clique set that allows us to identify partial covers in the graph that can be associated with dual feasible integer solutions to the linear programing equivalent of (NC). We then give a labeling procedure which successively modifies these partial covers so as to cover more and more edges of the graph, thus representing less and less infeasible solutions to (NC), without giving up the dual feasibility property in relation to the above mentioned linear program. Since the procedure in general stops short of finding a primal feasible (and thereby optimal) solution, some enumeration is necessary and we describe how this can be carried out within the same framework. Namely, when a problem is partitioned, the resulting dual node-clique sets need to be only locally modified to serve the same purpose in the graphs corresponding to the sub-problems created by the partition. The algorithmic applications of these

ideas are described in detail, and finally some initial computational experience and some ideas on extensions of our algorithm are outlined.

2. "DUAL" NODE-CLIQUE SETS

A set $\tilde{V}$ of nodes in a graph $G=(V, E)$ is called a cover if every edge of G is incident with some node in $\tilde{V}$. A cover is minimum if there exists no cover of smaller cardinality.

We will say that $\tilde{V}\subset V$ is a minimum partial cover if $\tilde{V}$ is a minimum cover in the subgraph G' obtained from G by removing all edges not incident with any node in $\tilde{V}$.

A set $Q\subseteq V$ of pairwise adjacent nodes will be called a clique.

As any other linear integer program, (NC) can be restated as a linear program, whose constraints are (or include) the facets of the convex hull H_I of integer solutions to (LNC). Some of these facets, as mentioned, are associated with the maximal cliques of G; and since the inequalities of (LNC) are themselves associated with edges, which can be viewed as 2-cliques, (NC) can be restated as the linear program

$$\text{Min} \sum_{j\in V} y(j)$$

$$\text{(LPNC)} \qquad \text{s.t.} \sum_{j\in Q_i} y(j)\geq |Q_i|-1,\ \forall Q_i\in K$$

$$\sum_{j\in V} d_{hj}y(j)\geq d_{ho},\ h\in F$$

$$y(j)\geq 0,\ j\in V$$

Here $y(j)$ denotes the variable associated with node j, while K is the family of all cliques of G; hence the set of inequalities indexed by K includes all the facets of H_I associated with maximal cliques, whereas the inequalities indexed by F are all those (not explicitly given) facets of H_I not associated with cliques.

The dual of (LPNC) is the linear program

$$\text{Max} \sum_{Q_i\in K} (|Q_i|-1)v(Q_i)+\sum_{h\in F} d_{ho}u_h$$

$$\text{(LNDC)} \qquad \text{s.t.} \sum_{Q_i|j\in Q_i} v(Q_i)+\sum_{h\in F} d_{hj}u_h\leq 1,\ j\in V$$

$$v(Q_i)\geq 0,\ u_h\geq 0,\ \forall Q_i\in K,\ h\in F$$

Note that this problem is a relaxation of (EM), in the sense that a feasible solution to (LDNC) can be associated with every feasible solution to (EM). Indeed, the variables of (EM) are associated with the edges, hence the 2-cliques, of G; therefore, they are among the variables of (LDNC) associated with the cliques of G. Further, there ia a 1—1 correspondence between the inequalities of (EM) and those of (LDNC), such that each variable of (EM) has the same coefficient in the j-th inequality of (EM) as in the j-th inequality of (LDNC).

For any node set $\tilde{V}\subseteq V$, the vector $y\in R^p$ defined by

$$y(j)=\begin{cases}1 & j\in\tilde{V}\\ 0 & j\in V-\tilde{V}\end{cases}$$

will be called the solution defined by $\tilde{V}$, and denoted $y(\tilde{V})$. A solution $y(\tilde{V})$ will be termed dual feasible if $[y(\tilde{V}), s]$, where s is the vector of slack variables whose value is uniquely determined by $y(\tilde{V})$, is a basic dual feasible solution to (LPNC) restated in equality form.

The following theorem gives a sufficient condition for a node set $\tilde{V}$ to be a minimum partial cover.

THEOREM 1: $\tilde{V} \subset V$ is a minimum partial cover, if there exists a set $\tilde{K}$ of cliques $Q_i \subset V$, $i=1, \ldots, t$, such that

(i) $Q_i \in \tilde{K}$, $Q_j \in \tilde{K}$, $i \neq j \Rightarrow Q_i \cap Q_j = \emptyset$

(ii) $j \in \tilde{V} \Rightarrow j \in Q_i$ for some $Q_i \in \tilde{K}$

(iii) $|Q_i \setminus \tilde{V}| = 1$ for all $Q_i \in \tilde{K}$.

PROOF: We will show that $y(\tilde{V})$ is dual feasible, hence optimal for the problem defined by those constraints that it satisfies. The vector $(\bar{v}, \bar{u})$, defined by $\bar{u}_h = 0$, $\forall h \in F$, and

$$\bar{v}(Q_i) = \begin{cases} 1 & Q_i \in \tilde{K} \\ 0 & Q_i \in K - \tilde{K}, \end{cases}$$

satisfies the constraints of (LDNC), since the cliques $Q_i \in \tilde{K}$ are disjoint and thus the left hand side of each inequality is at most 1.

Further, from (ii) we have

$$y(j) > 0 \Rightarrow \sum_{Q_i | j \in Q_i} v(Q_i) + \sum_{h \in F} d_{hj} \bar{u}_h = 1,$$

and from (iii)

$$\sum_{j \in Q_i} y(j) > |Q_i| - 1 \Rightarrow Q_i \notin \tilde{K}$$
$$\Rightarrow \bar{v}(Q_i) = 0$$

i.e., complementary slackness holds for the pair of solutions $y(\tilde{V})$ and $(\bar{v}, \bar{u})$. Hence $y(\tilde{V})$ is dual feasible. Q.E.D.

The algorithm to be discussed below generates a sequence of minimum partial covers $\tilde{V}$, and associated sets $\tilde{K}$ of cliques, satisfying requirements (i), (ii), (iii) of Theorem 1. Since each pari $(\tilde{V}, \tilde{K})$ defines, as shown above, solutions to a dual pair of linear programs, we will call $(\tilde{V}, \tilde{K})$ a *dual node-clique set*. The cliques in $\tilde{K}$ will be distinguished from each other and from those in $K - \tilde{K}$ by labeling. Thus a dual node-clique set $(\tilde{V}, \tilde{K})$ consists of a minimum partial cover $\tilde{V}$ and a collection $\tilde{K}$ of labeled cliques. These dual node-clique sets play a central role in the procedure to follow.

Our algorithm is an enumerative procedure consisting of the following major steps, to be described in detail in the following sections.

Finding an Initial Solution

A starting dual node-clique set $(\tilde{V}, \tilde{K})$ is generated by a one-pass labeling procedure.

Reducing Infeasibility

Given a dual node-clique set $(\tilde{V}, \tilde{K})$ for the current subproblem, primal infeasibility of the solution $y(\tilde{V})$ is gradually reduced, while integrality and dual feasibility are preserved. This is accomplished by a labeling procedure which successively updates the dual node-clique set $(\tilde{V}, \tilde{K})$, and whose steps correspond to sequences of dual simplex pivots in (LPNC). The number of steps in this procedure is bounded by $p(q+1)$.

If the solution becomes feasible then it is optimal for the current subproblem. Another subproblem is selected and step 1 is repeated.

If the infeasibility-reducing procedure ends before the current subproblem becomes feasible, then the latter is partitioned.

Branching and Reestablishing Dual Feasibility

The current subproblem is split into two new subproblems, one in which a certain node j is forced into the cover, and another one in which node j is forced out of the cover, while all nodes adjacent to j are forced into the cover. The nodes that are forced (into or out of the cover) are removed from the graph along with the edges incident with them, so that each new subproblem is associated with a proper subgraph of the graph of the parent problem.

Dual node-clique sets $(\tilde{V}, \tilde{K})$ are determined for each new subproblem from the dual node clique set of the parent problem. If any of the two new subgraphs is bipartite, the corresponding subproblem is solved as an assignment problem.

Bounding and Subproblem Selection

For each new subproblem created by branching, a lower bound on the objective function value is available from the dual feasible solution $y(\tilde{V})$. Another bound, often sharper, is generated by solving (LNC), which is accomplished by solving an associated assignment problem. The subproblem with the smallest lower bound is then selected for processing.

With a proper selection rule, the depth of the search-tree is bounded by the cardinality of a minimum node cover.

3. FINDING AN INITIAL SOLUTION

Notice that $(\tilde{V}_M, \tilde{K}_M)$, where $\tilde{K}_M$ is the set of 2-cliques (pairs of nodes) specified by a matching M in G, and $\tilde{V}J$ contains one of the two nodes of each such pair, satisfies the requirements (i), (ii), (iii) of Theorem 1 and hence defines a dual feasible solution to (LPNC).

Thus, the starting point for our procedure is an edge matching M in G which gives rise to a minimum partial cover $\tilde{V}_M$ of the same cardinality.

Of course, the higher the cardinality of M and thereby the value of the initial dual feasible solution $y(\tilde{V}_M)$, the better it is. A natural choice for M would therefore be a maximum matching i.e., an optimal solution to (EM). The fact that there is a polynomially bounded algorithm for that problem also seems to support this view.

We found however, that very good starting solutions could be obtained by means of a simplified procedure that finds a reasonably good matching in just one pass through the graph. The procedure, which is similar in spirit to Edmonds' algorithm, is the following:

(a) Starting from any node in the graph, attempt to partition V into two sets by alternatingly putting adjacent nodes into different sets.

(b) If (a) is interrupted by an inconsistency, an odd cycle C_i has been located and can be identified by tracing back along the paths by which the latest node was reached. The graph is then reduced by shrinking C_i, i.e., replacing it by a single node adjacent to all nodes $j \notin C_i$ adjacent to C_i.

(c) When the node set of the reduced graph has thus been partitioned, a maximum matching is found in the reduced (bipartite) graph.

(d) Odd cycles are successively unshrunk and k_i edges of each cycle C_i are added to the matching, where $k_i = \frac{|C_i|-1}{2}$.

PROPOSITION: The above procedure is consistent and finds a matching in G.

PROOF: The steps (a)–(c) can always be carried out since a graph is bipartite if and only if it has no odd cycles. To prove the validity of (d) we use an inductive argument. Suppose that at a given stage of the unshrinking step we have a matching M' in the current graph G' and we want to unshrink node n_{C_i} that represents an odd cycle C_i. Since M' is a matching in G', at most one edge of M' is incident with n_{C_i}. Let (j, n_{C_i}) be this edge. Then the graph G'' obtained from G' by unshrinking n_{C_i}, has an edge (j, h) for some $h \in C_i$.

Pick any such edge to be put, along with the edges of M', in the new matching M'' to be defined in G''. This will leave exactly $2k_i$ nodes of C_i exposed relative to M', which allows us to place k_i matching edges of C_i into M''. Since we can always find a matching in a bipartite graph and the above procedure can be reapplied whenever a node is unshrunk, the proof is complete.

To illustrate the portion of the algorithm that has been described so far consider the graph shown in Figure 1. The crossed edges constitute the feasible matching M found by our procedure, which in this case also happens to be maximum, as is easily verified. The pairs of nodes defined by M constitute the initial set $\tilde{K}$ of labeled cliques. The crossed nodes are in the corresponding minimum partial cover $\tilde{V}_M$.

We see that two edges (the circled ones) are not covered and the associated solution $y(\tilde{V})$, while dual feasible, is therefore not primal feasible. We now turn to the method for improving a minimal partial cover by introducing into $\tilde{K}$ cliques of cardinality greater than 2.

4. REDUCING INFEASIBILITY

The following procedure is uesd to reduce the (primal) infeasibility of a given dual feasible solution, while preserving integrality and dual feasibility. Each of the steps 2(a), 2(b), or 2(c) below corresponds to one or several dual simplex pivots.

1. Scan E for edges that are not covered by $\tilde{V}$, i.e., for which the corresponding constraint in (LPNC) is not satisfied.

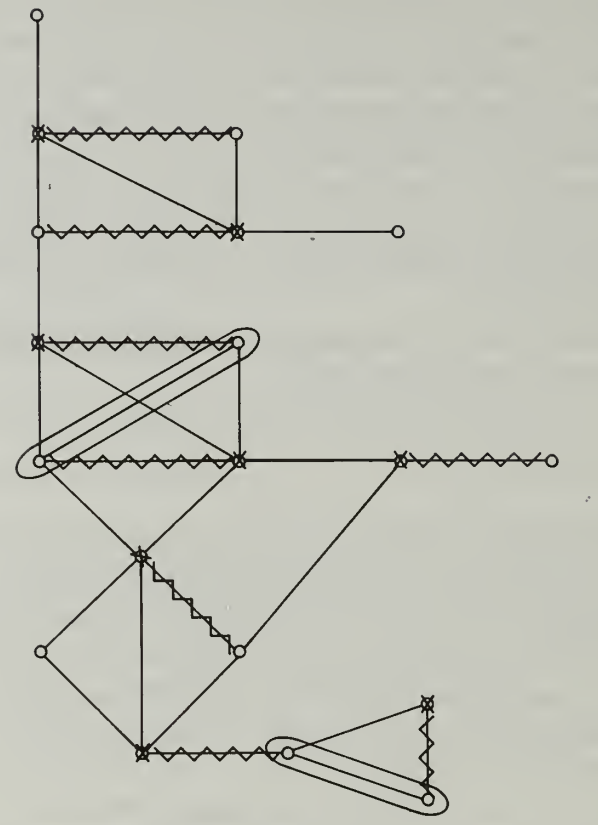


FIGURE 1

2. For each edge (i, j) that is not covered, attempt to perform one of the following steps in turn:

(a) (First labeling step)

If neither i nor j belongs to a labeled clique, label the 2-clique $\{i, j\}$. Then put into $\tilde{V}$ either i or j, whichever covers more edges not yet covered.

(b) (Reassignment step)

If either i or j can replace in $\tilde{V}$ one of the members of the labeled clique to which it belongs without creating any infeasibilities, make the switch to cover (i, j)

(c) (Second labeling step)

Find a largest unlabeled clique Q_*, if it exists, such that

(i) $i, j \epsilon Q_*$ and $|Q_*| \geq 3$

(ii) $Q_h \cap Q_* = \emptyset$ for all $Q_h \epsilon \tilde{K}$ such that $Q_h \not\subset O_*$ and $|Q_h| \geq 3$

(iii) Q_* contains a labeled clique or a node not belonging to a labeled clique

Then proceed as follows:

(α) Label Q_*

(β) Put into $\tilde{V}$ all but one of the nodes in Q_*

(γ) Delete from $\tilde{V}$ all nodes $j \epsilon \tilde{V} \sim Q_*$ belonging to labeled 2-cliques incident with Q_*

(δ) Delete from $\tilde{K}$ ("unlabel") all labeled cliques contained in Q_* and all labeled 2-cliques incident with Q_*

To continue the above example, we see that one of the uncovered edges is in a 4-clique Q_α. If Q_α is labeled we can bring the node n_α into $\tilde{V}$ and get the situation shown in Figure 2.

We are left with one uncovered edge and to eliminate it we label the triangle Q_β and put n_β into $\tilde{V}$ (see Figure 3). This however forces n_γ out of the solution and gives rise to two new uncovered edges.

Labeling any of the triangles on which n_γ lies makes it possible to reintroduce n_γ into the cover. We then get the dual node-clique set $(\tilde{V}, \tilde{K})$ shown in Figure 4, where $\tilde{K}$ contains the circled cliques, plus the 2-cliques corresponding to the crossed edges. This solution happens to be (primal) feasible and is thus also optimal, which means that the crossed nodes in Figure 4 constitute a minimum node cover.

Next we prove the following property of the above procedure. Let

$$E(y)=\{(i,j)\epsilon E|y(i)+y(j)\geq 1\}.$$

THEOREM 2: The infeasibility-reducing procedure produces a sequence of dual node-clique sets $(\tilde{V}_i, \tilde{K}_i)$, and associated dual feasible integer solutions $y(\tilde{V}_i)=y^i$, $i=1, \ldots, r$, such that

(α) $e_v y^i \geq e_v y^{i-1}$ $i=2, \ldots, r$

(β) y^i satisfies all constraints corresponding to cliques $Q\subset\tilde{K}^{i-1}$, $|Q|\geq 3$, satisfied by y^{i-1}

(γ) If y^i is obtained by step (a) or (b), then $|E(y^i)|>|E(y^{i-1})|$; and in the case of step (a) $e_v y^i > e_v y^{i-1}$

(δ) If y^i is obtained by step (c), then y^i satisfies a constraint corresponding to a clique $Q\epsilon\tilde{K}_i$, $|Q|\geq 3$, which is violated by y^{i-1}

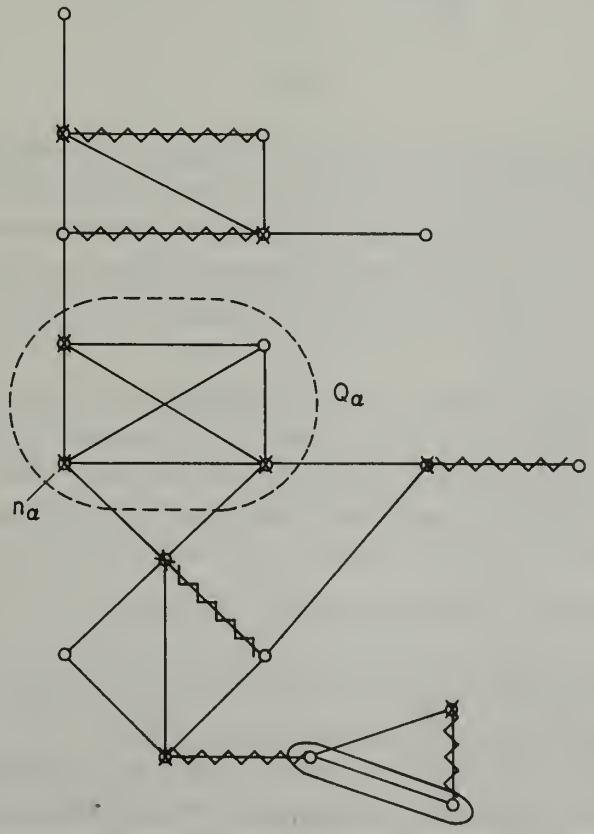


FIGURE 2

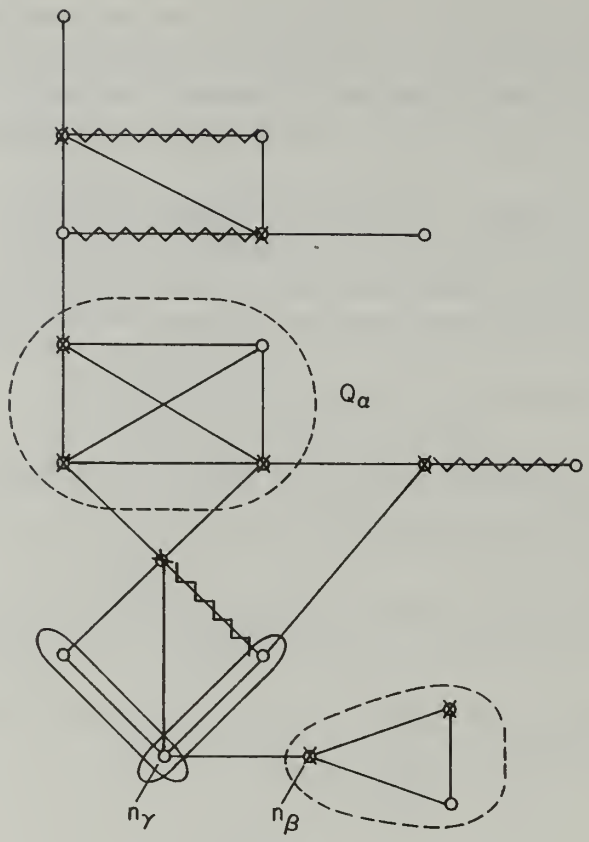


FIGURE 3

PROOF: We first prove that the conditions of Theorem 1 are preserved by the procedure.

(i) Step 2(a) labels a 2-clique whose nodes do not belong to any clique; step 2(b) does not label any clique; and step 2(c) labels a clique Q_* only if Q_* is disjoint from all labeled cliques Q such that $|Q| \geq 3$, while it "unlabels" (removes from $\tilde{K}$) all labeled cliques which have a nonempty intersection with Q_*. Thus each of the three steps leaves all the labeled cliques pairwise disjoint, which is condition (i) of Theorem 1.

(ii) If a node is put into $\tilde{V}$ under any of the steps 2(a), 2(b), or 2(c), it belongs to some labeled clique. Conversely, if a labeled clique Q is "unlabeled," which can only happen in step 2(c), then Q is either contained in another labeled clique Q_*, or is a 2-clique such that $Q \cap Q_* \neq \emptyset$. In the latter case, if the node $Q \sim Q_*$ belongs to $\tilde{V}$, it is taken out of $\tilde{V}$. Hence, the set $\tilde{V}$ resulting from any of the three steps, is contained in $\bigcup_{i \in \tilde{K}} Q_i$, which is condition (ii) of Theorem 1.

(iii) Step 2(a) labels a 2-clique Q and puts into $\tilde{V}$ one of the two nodes of Q, thus making sure that $|Q \cap \tilde{V}| = 1$. Step 2(b) replaces a node $j \in Q$ of $\tilde{V}$ by another node $h \in Q$, which leaves $|Q \cap \tilde{V}|$ unchanged. Step 2(c) labels a new clique Q_* and puts into $\tilde{V}$ all but one of the nodes of Q_*. Hence each step preserves property (iii) of the node-clique set $(\tilde{V}, \tilde{K})$.

Thus, from Theorem 1, the above procedure generates a sequence of dual node-clique sets $(\tilde{V}_t, \tilde{K}_t)$, and associated dual feasible integer solutions y^t. To show that the sequence has the other properties claimed in Theorem 2, we examine each of the steps 2(a), 2(b), 2(c) in turn.

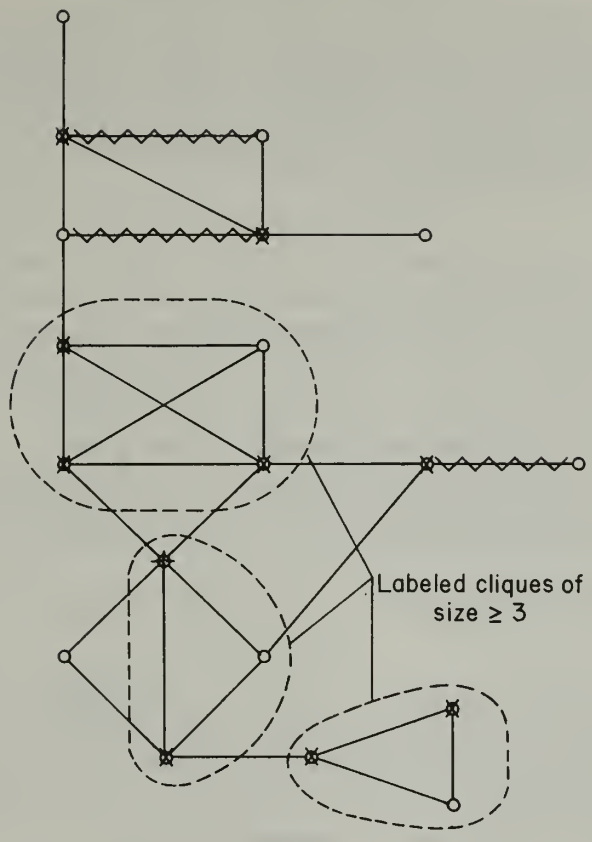


FIGURE 4

Step 2(a). Under this step $\tilde{V}_t$ is replaced by $\tilde{V}_{t+1}=\tilde{V}_t \cup \{j\}$ for some $j \epsilon V-\tilde{V}_t$; thus (α), (β) and (γ) clearly hold.

Step 2(b). Under this step $\tilde{V}_t$ is replaced by $\tilde{V}_{t+1}=(\tilde{V}_t-\{j\}) \cup \{h\}$ for some pair $\{j, h\}$ belonging to the same labeled clique Q; hence (α) holds after the step. Also, since j and h do not belong to any other labeled clique but Q, (β) holds. Finally, the pair $\{j, h\}$ is chosen so as to increase by one the number of edges covered by $\tilde{V}_t$, i.e., (γ) also holds.

Step 2(c). Under this step, $\tilde{V}_t$ is replaced by $\tilde{V}_{t+1}=(\tilde{V}_t-S) \cup S'$, where S and S' are the subsets of nodes deleted from and introduced into $\tilde{V}$ under steps 2(c)(α) and 2(c)(β) respectively.

We claim that $|S'| \geq |S|$. From 2(c)(α), each node of S belongs to a two-clique which intersects Q_* without being contained in it. For each such two-clique $\{i, j\}$, where $i \epsilon S$, we have $j \epsilon Q_* \sim \tilde{V}_t$. Let S'' be the set of these nodes, i.e.,

$$S''=\{j \epsilon Q_* \sim \tilde{V}_t | \{i, j\} \epsilon \tilde{K}_t,\ i \epsilon S\}.$$

From what we just said, $|S''|=|S|$, i.e., Q_* contains at least $|S|$ nodes belonging to two-cliques of the above type and not belonging to $\tilde{V}_t$. By property (iii), Q_* also contains at least one additional node not belonging to $\tilde{V}_t$, which raises the number of nodes in $Q_* \sim \tilde{V}_t$ to at least $|S|+1$. Since by step 2(c)(β) all but one of the nodes in $Q_* \sim \tilde{V}_t$ are introduced into $\tilde{V}_{t+1}$, it follows that the number $|S'|$ of such nodes is at least equal to $|S|$. This proves the claim.

From $|S'| \geq |S|$, it follows that property (α) of Theorem 2 holds after the step 2(c). Properties (β) and (δ) also hold, since no vertex belonging to a clique of cardinality $k \geq 3$ is deleted from $\tilde{V}_i$, and $\tilde{V}_{i+1}$ contains all but one of the vertices of the newly labeled clique Q_*, where $|Q_*| \geq 3$ and $|\tilde{V}_i \cap Q_*| \leq |Q_*| - 2$, i.e., the constraint of (LPNC) corresponding to Q_* is satisfied by y^{i+1}, but not by y^i. The last inequality follows from the fact that if (i, j) is the edge found under step 1, then $\{i, j\} \epsilon Q_* \sim \tilde{V}_i$. Q.E.D.

COROLLARY 2.1: After at most $p(q+1)$ steps, of which at most p are of type 2(c), the infeasibility-reducing procedure either finds an optimal solution or cannot be continued.

PROOF: Each time step 2(c) is applied, the number of nodes belonging to labeled cliques of cardinality $k \geq 3$ is increased by at least one, while the other steps leave this number unchanged. Hence $|V| = p$ is an upper bound on the total number of steps 2(c).

Each of steps 2(a) and 2(b) increases by one the number of edges covered; hence $|E| = q$ is an upper bound on the number of steps 2(a) and 2(b) between any two consecutive steps 2(c). Since the total number of steps 2(c) is bounded by p, the total number of steps 2(a) and 2(b) is bounded by $p \times q$.

Summing up the above, $p \times q + p = p(q+1)$ is an upper bound on the total number of steps in the procedure. Q.E.D.

5. BRANCHING AND REESTABLISHING DUAL FEASIBILITY

Although in the above example the infeasibility-reducing procedure found the optimal solution directly, this cannot be expected in general, as illustrated by the example shown in Figure 4a. Assuming that the 4-clique is labeled, the procedure will stop short of removing the infeasibilities at the edges e_1, e_2.

In such a case we branch, i.e., partition the current subproblem.

Our partitioning rule is the usual one, i.e., we create two subproblems by setting a variable in (LPNC) to 0 and 1 respectively. However, due to the special structure of the problem [i.e., the presence of constraints of the form $y(i) + y(j) \geq 1$, $(i, j) \epsilon E$], $y(j) = 0$ implies $y(i) = 1$ for all i adjacent to j, and thus the partitioning rule becomes

$$\{y(j) = 1\} \vee \{y(j) = 0,\ y(i) = 1,\ \forall i\colon (i, j) \epsilon E\}.$$

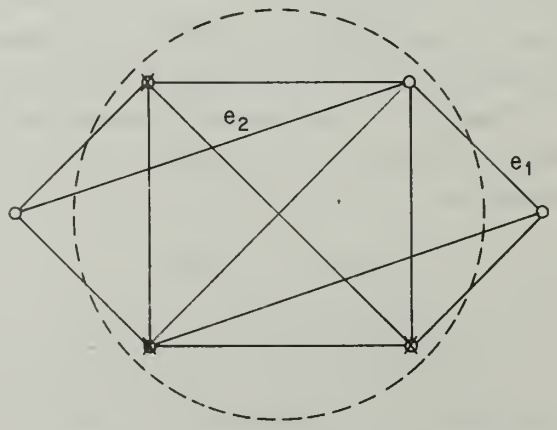


FIGURE 4a

For both subproblems created by the partition the graph can be reduced accordingly.

Since the current solution may not be dual feasible for the subproblems, it is necessary to re-establish this property. It is one of the main advantages of our approach that the all integer dual feasible solution associated with the original graph can be made valid for the subproblems, by only local modifications.

The partitioning procedure can then be stated as follows.

Let $A(j)=\{i\epsilon V|(i, j)\epsilon E\}$.

0. Choose $i\epsilon V-\tilde{V}$ such that

$$|A(i)\backslash\tilde{V}|=\max_{j\epsilon V-\tilde{V}}|A(j)\backslash\tilde{V}|$$

1. Partition the current problem by

$$\{y(i)=1\}\vee\{y(i)=0,\ y(j)=1,\ \forall j\epsilon A(i)\}.$$

This creates two subproblems and associated subgraphs, as follows.

SUBPROBLEM 1 is (LPNC) defined on the graph G_1, obtained from G by removing node i and all edges (i, j), $j\epsilon V$.

SUBPROBLEM 2 is (LPNC) defined on the graph G_2, obtained from G by removing nodes i and $j\epsilon A(i)$, and all edges (i, j) and (j, h), $j\epsilon A(i)$, $h\epsilon V$.

2. Redefine $(\tilde{V}, \tilde{K})$ as follows:

Let $(\tilde{V}^{\text{old}}, \tilde{K}^{\text{old}})$ and $(\tilde{V}^{\text{new}}, \tilde{K}^{\text{new}})$ denote the dual node-clique sets of the parent-problem and (any) one of the subproblems, respectively. Let V^{new} be the node set of the new graph.

Denote

$$I=\left\{i\epsilon V-\tilde{V}^{\text{old}}\middle|\begin{array}{l}y(i)\text{ is fixed at 1,}\\ i\epsilon Q\text{ for some }Q\epsilon\tilde{K}^{\text{old}}\end{array}\right\}$$

For each $i\epsilon I$, choose a node $j(i)$ $\epsilon Q\cap V^{\text{new}}$ having as few adjacent nodes in $V^{\text{new}}\sim\tilde{V}^{\text{old}}$ as possible, and let $J=\{j(i)|i\epsilon I\}$. Then define

$$\tilde{V}^{\text{new}}=V^{\text{new}}\cap\tilde{V}^{\text{old}}-J$$

Further, let

$$S=\{Q\epsilon\tilde{K}^{\text{old}}|Q\not\subseteq V^{\text{new}}\}$$

and

$$S'=\{Q\cap V^{\text{new}}|Q\epsilon S,\ |Q\cap V^{\text{new}}|\geq 2\}.$$

Then define

$$\tilde{K}^{\text{new}}=\{\tilde{K}^{\text{old}}-S\}\cup S'$$

To illustrate the above on a graph that requires some enumeration, consider the example shown in Figure 5. The maximum matching and the corresponding partial cover are crossed. There is an infeasibility in the 4-clique, but that is immediately eliminated by labeling it. The remaining infeasibility cannot be removed though (Figure 6). (Note that the infeasible edge e_A in figure 6 is adjacent to an odd cycle that is not a clique.)

The procedure specifies n_α as the branching node. Subproblem 2 is given by

$$y(n_\alpha)=0\text{ and }y(i)=1\text{ for }i=n_\beta, n_\gamma, n_\delta.$$

This leaves the graph shown in Figure 7. We have feasibility and thus an optimal cover for this subgraph, which, together with the nodes forced into the cover by the branching step, gives a cover of cardinality 9 for the initial graph G.

Subproblem 1 is defined by $y(n_\alpha)=1$, and shown in Figure 8. Note that n_δ had to be taken out of $\tilde{V}$ to preserve complementarity, since it belonged to the same labeled 2-clique as n_α.

Now n_δ is in a clique, but one which cannot be labeled, since it already contains a node from a labeled clique. So we branch on n_δ. For Subproblem 2, defined by $y(n_\delta)=1$, we get a solution that is immediately found to be feasible. Subproblem 2, defined by $y(n_\delta)=0$ and $y(i)=1$ for all i adjacent to n_δ, is shown in Figure 9.

The labeled 4-clique is now reduced to a triangle. The infeasibility from forcing n_θ out of $\tilde{V}$ is immediately corrected by labeling the triangle containing n_θ and we have an optimal solution to the subproblem.

Thus we terminate with the search tree shown in Figure 10.

We have found and verified three alternative optimal node-covers of cardinality nine.

We now prove the following property of the procedure.

THEOREM 3: $(\tilde{V}^{\text{new}}, \tilde{K}^{\text{new}})$ is a dual node-clique set if $(\tilde{V}^{\text{old}}, \tilde{K}^{\text{old}})$ is.

PROOF: We show that conditions (i), (ii), (iii) of Theorem 1 are satisfied.

(i) This property is clearly preserved, since all new members of $\tilde{K}$ are obtained from old members by deletion of some nodes.

(ii) From the definition of $\tilde{K}^{\text{new}}$, the only nodes $j\epsilon\tilde{V}^{\text{new}}\cap\tilde{V}^{\text{old}}$ contained in some clique of $\tilde{K}^{\text{old}}$ but no clique of $\tilde{K}^{\text{new}}$ are those belonging to some 2-clique $\{i,j\}\epsilon\tilde{K}^{\text{old}}$, such that $i\epsilon V-V^{\text{new}}$. But since $j\epsilon\tilde{V}^{\text{old}}$, from property (iii) of $(\tilde{V}^{\text{old}}, \tilde{K}^{\text{old}})$ it follows that $i\epsilon V-\tilde{V}^{\text{old}}$. Further, the variable $y(i)$ must have been fixed at 1, for if it had been fixed at 0, then $y(j)$ would have been fixed at 1 and j would not be a node of the new graph. Hence $i\epsilon I$ and therefore $j\epsilon J$, where I and J are the sets used in the definition of $\tilde{V}^{\text{new}}$. Thus $j\epsilon V^{\text{new}}-\tilde{V}^{\text{new}}$, which proves that property (ii) holds for $(\tilde{V}^{\text{new}}, \tilde{K}^{\text{new}})$.

(iii) For any clique $Q\epsilon\tilde{K}^{\text{new}}\cap\tilde{K}^{\text{old}}$, from the definition of $\tilde{V}^{\text{new}}$ we have $Q\cap\tilde{V}^{\text{new}}=Q\cap\tilde{V}^{\text{old}}$, hence property (iii) is preserved.

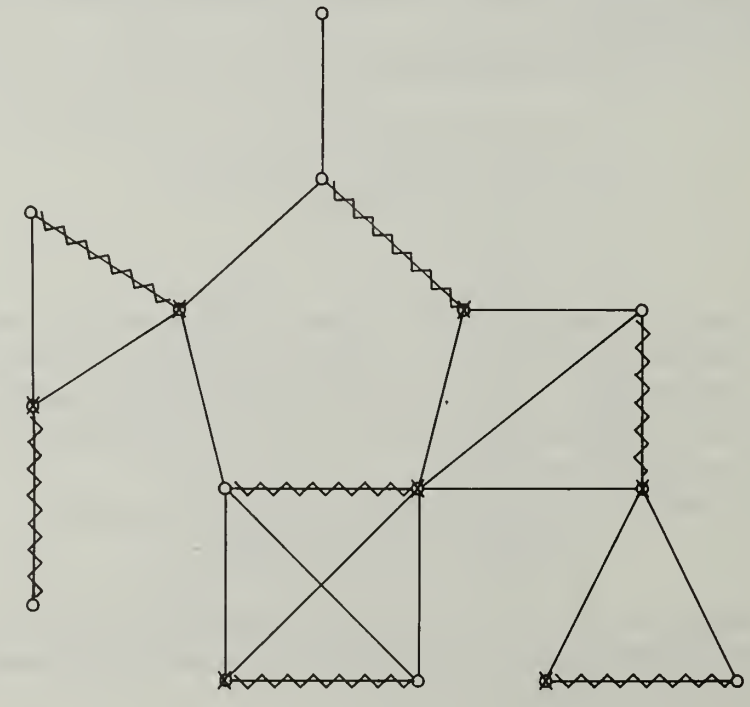


FIGURE 5

For each clique $Q \epsilon S' = \tilde{K}^{\text{new}} / \tilde{K}^{\text{old}}$, there exists some $Q' \epsilon \tilde{K}^{\text{old}}$ such that $Q \subset Q'$. Then $(Q' - Q) \cap (V - \tilde{V}^{\text{old}})$ is either empty, or consists of one element, say i. In the latter case, $i \epsilon I$ and an element $j(i)$ of Q is removed from $\tilde{V}^{\text{old}}$, according to the rule defining $\tilde{V}^{\text{new}}$; while in the former case, $Q \cap \tilde{V}^{\text{new}} = Q' \cap \tilde{V}^{\text{old}} \cap \tilde{V}^{\text{new}}$. In both cases, $|Q' \cap \tilde{V}^{\text{old}}| = |Q'| - 1$ implies $|Q \cap \tilde{V}^{\text{new}}| = |Q| - 1$. Q.E.D.

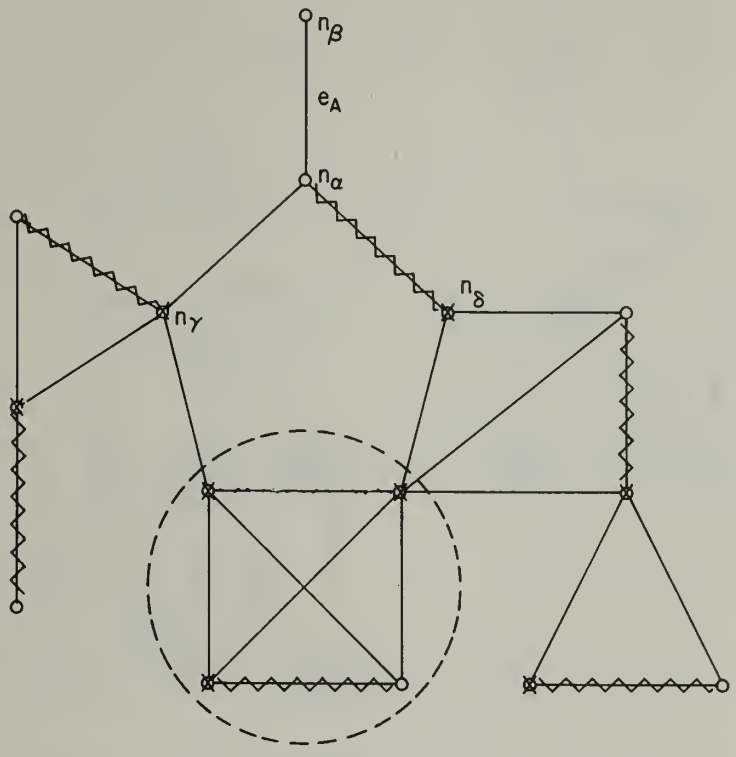


FIGURE 6

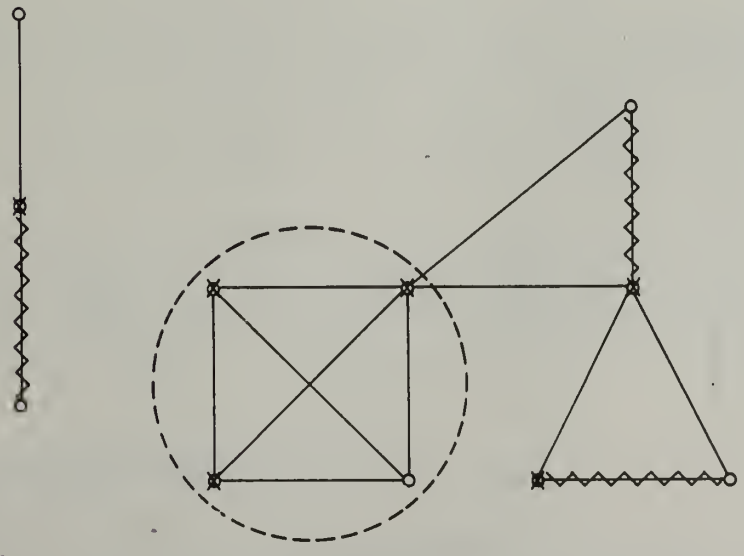


FIGURE 7

6. BOUNDING AND SUBPROBLEM SELECTION

Before any further work is done on a newly created subproblem, we check whether the associated graph is bipartite. If so, then the subproblem is solved as an assignment problem. Otherwise we turn to the task of calculating bounds.

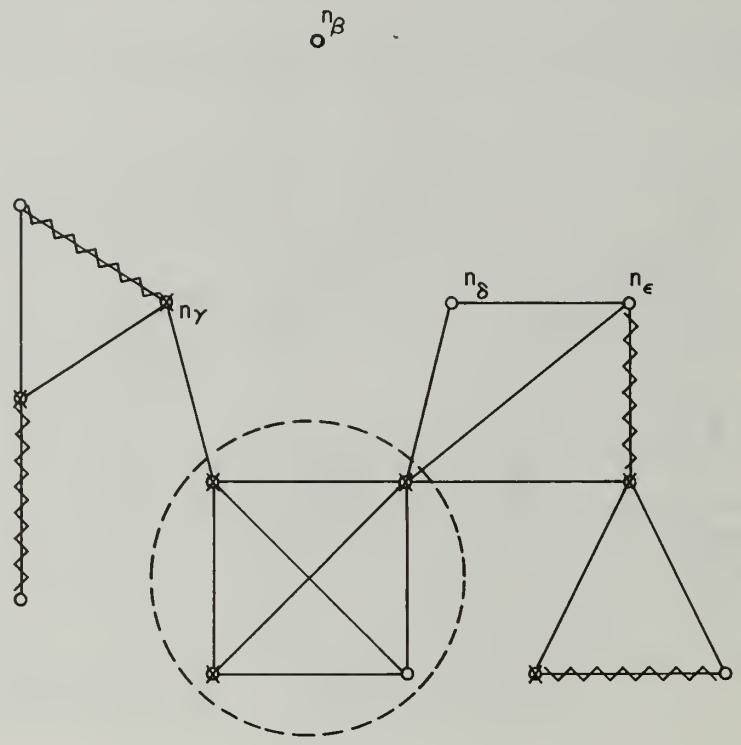


FIGURE 8

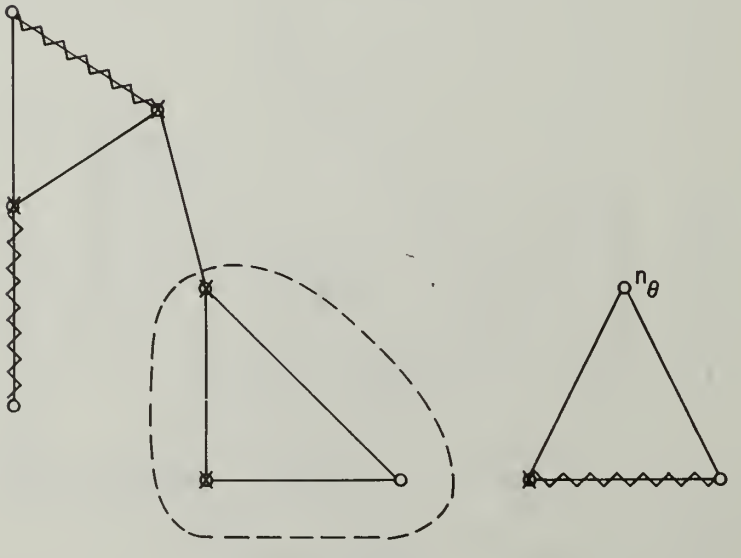


FIGURE 9

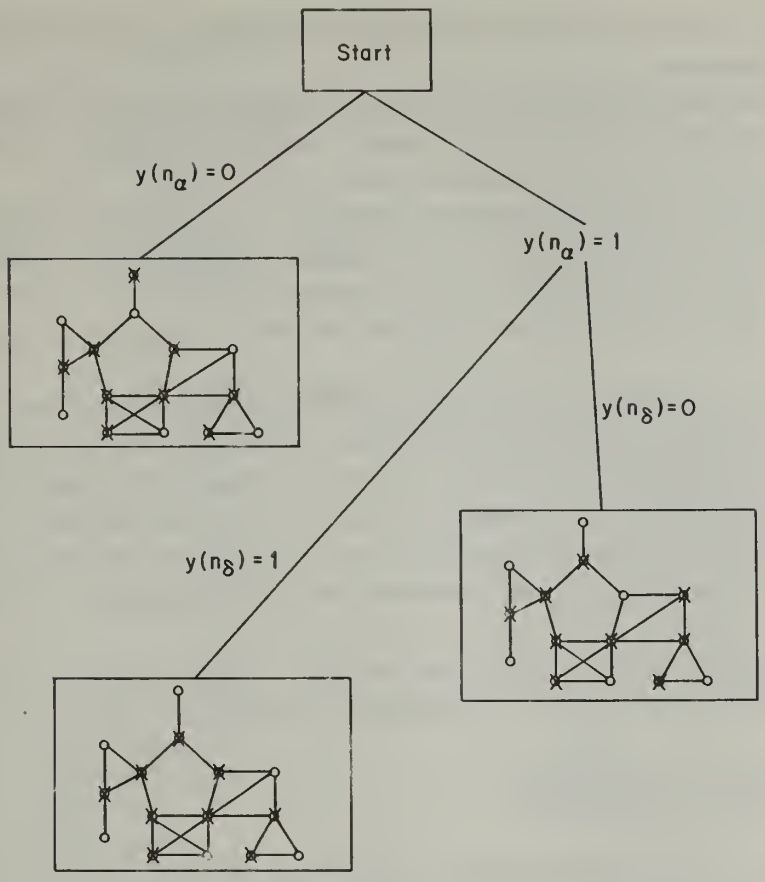


FIGURE 10

For all the subproblems created in the enumeration process, a lower bound on the optimal objective value is available from the dual feasible all-integer solution.

Another such bound is given by the solution to (LNC), and our experience indicates that the latter is often, though not always, sharper. To facilitate fathoming by comparison with the currently best feasible solution, we found it worthwhile to solve (LNC) for each subproblem. This is rather inexpensive, thanks to the following result attributed by Trotter [10] to Edmonds and Pulleyblank, which makes (LCN) equivalent to an assignment problem in a bipartite graph twice the size of G.

THEOREM 4: (Edmonds and Pulleyblank) Define a bipartite graph $G'=(V \cup V', E')$ where V' is a copy of V, i.e., $i' \epsilon V'$ if $i \epsilon V$, and $E'=\{(i, j') | i \epsilon V, j' \epsilon V, (i, j) \epsilon E\}$. Let (x, x'), where x and x' are 0–1 vectors having one component for each node of V and V' respectively, be an optimal solution to (NC) defined on G'. Then

$$\frac{x+x'}{2}$$

is an optimal solution to (LNC) defined on G.

PROOF: Obvious (see [10]).

In case the objective value found for (LNC) is fractional, the bound can of course be rounded upwards. In principle one could apply to fractional solutions the usual penalties used in branch and

bound. It turns out, however, that due to the high degeneracy of assignment problems, nothing is usually gained that way.

Each time two new subproblems are created, attempts are made to fathom both. If this is not possible, one of them is stored and the other is partitioned again. In this choice, preference is given to the subproblem where a variable is fixed at 0.

When both of the current subproblems are fathomed, a problem has to be selected from storage. In this choice the algorithm is guided by the tightest of the two lower boards just mentioned, and the problem with the lowest such bound is considered next.

To conclude the description of our algorithm we give a flow chart of the entire procedure (Figure 11).

In the above version of the algorithm, after each branching step we start processing one of the two newly created subproblems. We found this preferable to choosing the subproblem with the lowest bound after each branching step, since it keeps storage requirements low.

If, however, the other rule is preferred, then one can state the following.

THEOREM 5: If the subproblem with the lowest bound is selected for processing after each branching step, then the cardinality of a minimum node cover is an upper bound on the depth of the search tree.

PROOF: At each branching step, at least one new variable is fixed at 1 in each of the two subproblems created (if $y(i)$ is chosen for branching, $A(i) \neq \phi$). Q.E.D.

7. COMPUTATIONAL EXPERIENCE

The version of our algorithm illustrated in Figure 11 was programmed in Fortran IV for an 1108 Univac computer. Since the programming was undertaken mainly to test the validity of our approach, no particular care was taken to program the algorithm in an efficient manner at this stage. Instead, maximum flexibility in testing alternative procedures was emphasized and the results were evaluated primarily in terms of the number of iterations in the various parts of the routine.

As should be evident from the previous description of the algorithm, all calculations can be carried out by labeling in a computer representation of the graph. Thus we work essentially with:

(a) A list of size p, indicating for each node whether it is in $\tilde{V}$ and whether it belongs to a member of $\tilde{K}$, and if so, which one

(b) A list of edges

(c) An adjacency matrix

(d) A problem list. For each subproblem on the list it is only necessary to save (a), since (b) and (c) are basically the same for all of them. This enables one to keep fairly large problems entirely in core, although the very limited memory size of the 1108 (30 k words available and no possibility to use halfwords) has only allowed us to attempt problems with up to 50 nodes and 400 edges.

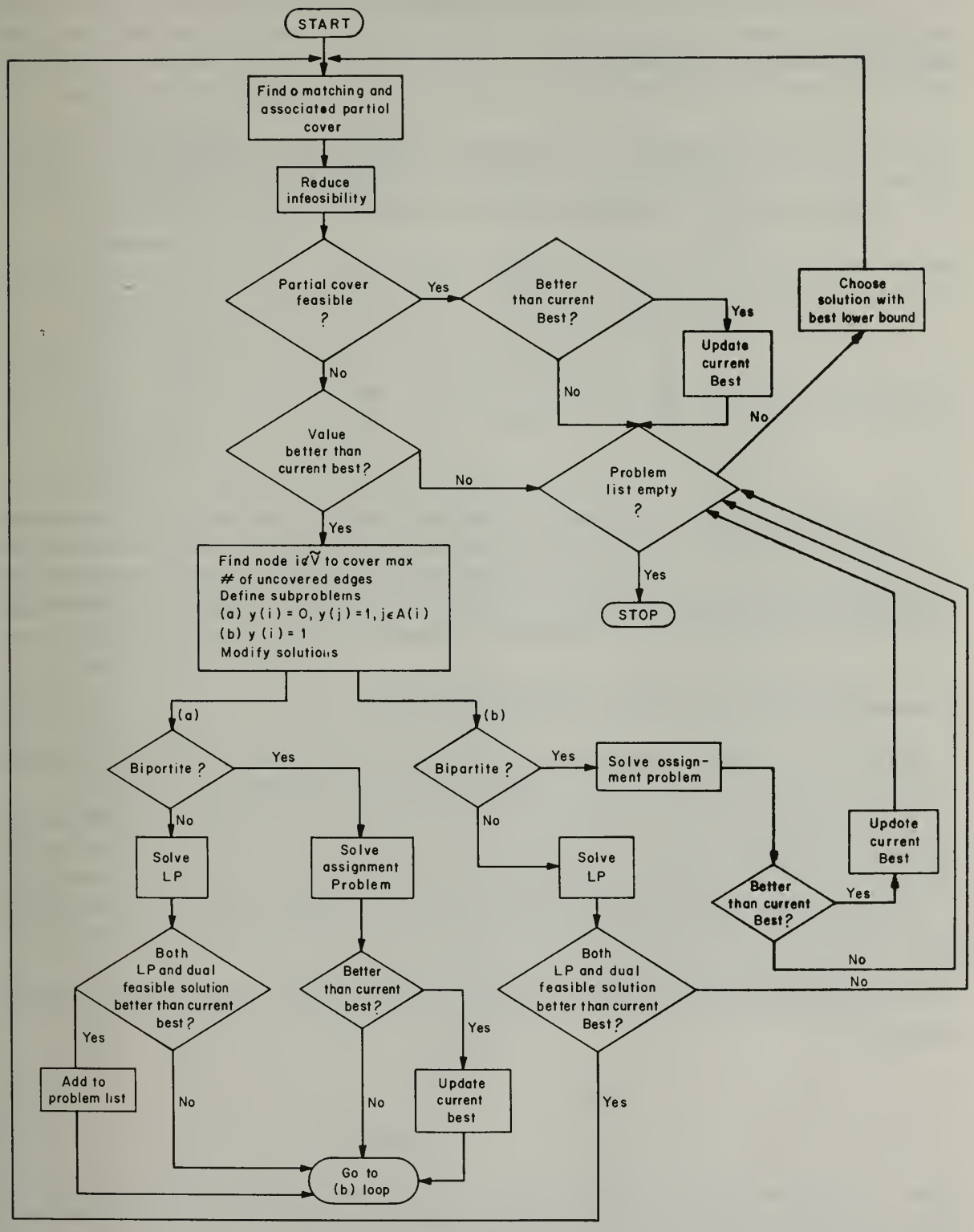

START
Find o matching and associated partiol cover
Reduce infeosibility
Partial cover feasible ?
Yes
Better than current Best?
Yes
Update current Best
Choose solution with best lower bound
No
No
No
Value better than current best?
No
Problem list empty ?
Yes
Yes
STOP
Find node i∈V̄ to cover max # of uncovered edges
Define subproblems
(a) y(i) = 0, y(j) = 1, j∈A(i)
(b) y(i) = 1
Modify solutions
(a)
(b)
Biportite ?
Yes
No
Solve LP
Solve assignment Problem
Bipartite ?
Yes
Solve assign-ment problem
No
Solve LP
Updote current Best
Better than current Best?
Yes
No
Both LP and dual feasible solution better than current best?
Yes
No
Better than current best?
Yes
No
Both LP and dual feasible solution better than current Best?
No
Yes
Add to problem list
Update current best
Go to (b) loop

FIGURE 11

A number of node covering problems randomly generated by Trotter [14] were kindly made available to us by him. The following tableau contains results on those of his problems that do not violate the above space requirements.

Problem	Number of nodes	Number of edges	Number of applications of partitioning routine	Number of applications of infeasibility-reducing routine
1	50	64	2	5
2	50	55	1	3
3	50	59	2	5
4	50	107	4	9
5	50	114	7	12
6	50	117	5	11
7	50	275	81	122
8	50	295	73	103
9	50	329	102	148

We see that the number of operations increases with the density of the graph (i.e., the number of arcs/number of nodes). In Trotter's experience [14] with these and other problems with 50 variables but higher density, the hardest problems are those of density approximately 25%, such as the three last ones in the above tableau. The use of the LP-bound decreases the number of iterations by approximately 50% to 75% and is also helpful in cutting down storage requirements.

8. POSSIBLE IMPROVEMENTS AND EXTENSIONS

As in the general set covering problem, there are a number of reductions that can be used to cut down the size of the graph before solving the node covering problem.

For instance, if the degree of a node exceeds an upper bound on the value of a minimum cover, then it must belong to any such cover. If a node is in a clique and only adjacent to other nodes in the clique, there is an optimal cover that does not contain it.

Features such as these were not incorporated in our present code which, as mentioned above, was mainly intended as a tool for testing the merits of our approach and for allowing us to experiment with different versions of the major components of the algorithm.

A more general problem than the one considered in this paper is the weighted node covering problem, i.e.,

$$\text{Min by} \cdot$$

$$(NC_b) \qquad A^T y \leq e_q$$

$$y \in \{0, 1\}^p.$$

where b is a p-vector of positive integers. We have recently found an extension of our algorithm to this problem, that employs essentially the same concepts and algorithmical steps, with appropriate modifications. This extension is left, however, to another paper.

BIBLIOGRAPHY

[1] Balas, E. and Samuelsson, "Finding a Minimum Node Cover in an Arbitrary Graph." Management Science Research Report, GSIA No 325, Carnegie-Mellon University (November 1973).

[2] Balas, E. and M. W. Padberg, "Set Partitioning: A Survey," SIAM Review, *18*, 710–760 (1976).

[3] Balas, E. and E. Zemel, "Graph Substitution and Set Packing Polytopes," Management Science Research Report, GSIA No. 384, Carnegie-Mellon University (January 1976), to appear, Networks.

[4] Balas, E. and E. Zemel, "Critical Cutsets of Graphs and Canonical Facets of Set Packing Polytopes," Management Science Research Report No. 385, Carnegie-Mellon University (February 1976); to appear, Mathematics of Operations Research.

[5] Berge, C., *Graphs and Hypergraphs* (North-Holland, 1973).

[6] Chvátal, V., "On Certain Polytopes Associated with Graphs," Centre de Recherches Mathématiques, CRM 238, Université de Montıéal (October 1972), Revised as CRM 397, (March 1974).

[7] Edmonds, J., "Maximum Matching and a Polyhedron with 0–1 Vertices," Journal of Research of the National Bureau of Standards 69B, 126–130 (1965).

[8] Gallai, T., "Über extreme Punkt und Kantenmengen," Annales Universitatis Scientorium Budapest, Eötvös, Le Sectio Mathematicn *2*, 133–138 (1959).

[9] Gavril, F., "Algorithms for Minimum Coloring, Maximum Clique, Minimum Covering by Cliques, and Maximum Independent Set of a Chordal Graph," SIAM Journal of Computation *1*, 180–187 (1972).

[10] Gavril, F., "Algorithms for a Maximum Clique and a Maximum Independent Set of the Circle Graph," Networks, *3*, 261–273 (1973).

[11] Nemhauser, G. and L. Trotter, "Vertex Packings: Structural Properties and Algorithms," Mathematical Programming, *8*, 232–248 (1975).

[12] Padberg, M., "On the Facial Structure of Set Packing Polyhedra," Mathematical Programming, *5*, 199–215 (1973).

[13] Shannon, C., "The Zero-Error Capacity of a Noisy Channel," IEEE, Transactions, *3*, 3 (1956).

[14] Trotter, L., "Solution Characteristics and Algorithms for the Vertex Packing Problem," Technical Report, 168, Department of Operations Research, Cornell University (January 1973).

[15] Trotter, L., "A Class of Facet Producing Graphs for Vertex Packing Polyhedra," Technical Report No. 78, Yale University (February 1974).

TWO-CHARACTERISTIC MARKOV-TYPE MANPOWER FLOW MODELS*

W. J. Hayne and K. T. Marshall

Naval Postgraduate School
Monterey, California

ABSTRACT

A two-dimensional state space Markov Model of a Manpower System with special structure is analyzed. Examples are given from the military services. The probabilistic properties are discussed in detail with emphasis on computation. The basic equations of manpower stocks and flows are analyzed.

I. INTRODUCTION

The simple fractional flow (or Markov-type) model of personnel movements through an organization has been widely analyzed (see for example Bartholomew [1], Blumen, Kogan and McCarthy [2], Lane and Andrew [5], Rowland and Sovereign [8] and has been widely applied, especially in military manpower planning (see U.S. Navy [9]). Other models such as the "cohort" and "chain" models (see Marshall [6] and Grinold and Marshall [3]) satisfy more realistic assumptions on personnel movement, but lack the convenient structure of the Markov model. The purpose of this paper is to present an extension of the Markov Model to one with a 2-dimensional state space. The state space is chosen so that the fractional flow matrix has a special structure which is then exploited.

In Section II the structure of the model is presented and in Section III examples are given. In Section IV we present the probabilistic properties of the model with emphasis on computationally tractable formulae. In Section V the structure of the model is exploited in the personnel stock and flow equations.

II. STRUCTURE OF THE FRACTIONAL FLOW MATRIX

We assume that for planning purposes an organization considers time in discrete periods, and that people are counted at the end of each period. When counted, a person is assumed to possess two characteristics i and j, and is said to be in state (i, j), where i represents the first characteristic (FC), $1 \leq i \leq n$, and j represents the second characteristic (SC), $l(i) \leq j \leq u(i)$. Here $l(i)$ and $u(i)$ are the lower and upper limits respectively for the SC when i is the FC. Also let $J(i) = \{j | l(i) \leq j \leq u(i)\}$, the set of SC's for FC i, and let w_i be the number of elements in $J(i)$.

* The work reported herein was supported by a grant from the U.S. Marine Corps.

Let $q_i(j, m)$, $j, m \in J(i)$ be the fraction of people in state (i, j) in a time period who move to state (i, m) in the next time period, and let Q_i be the $w_i \times w_i$ matrix $[q_i(j, m)]$. Let $p_i(j, m)$, $j \in J(i)$, $m \in J(i+1)$ be the fraction of people in state (i, j) in a time period who move to state $(i+1, m)$ in the next time period, and let P_i be the $w_i \times w_{i+1}$ matrix $[p_i(j, m]$. A basic assumption of our two-characteristic model is that movement in one time period from states with FC i can only be to states with FC i or $i+1$, or out of the system. Following the notation of earlier papers, let Q be the fractional flow matrix for all active states in the system. Then Q has the following structure:

$$Q = \begin{bmatrix} Q_1 & P_1 & & & \\ & Q_2 & P_2 & & \\ & & & & \\ & & & Q_{n-1} & P_{n-1} \\ & & & & Q_n \end{bmatrix} \tag{1}$$

where the zero matrices have been suppressed. Throughout this paper we assume that people can leave the system eventually from any state and thus $(I-Q)$ has an inverse, where I is the identity matrix. This inverse $(I-Q)^{-1}$ is called N, the fundamental matrix (see Kemeny and Snell [4]. For each i let $A_i = (I-Q_i-P_i)\bar{1}$, where $\bar{1}$ is a vector with every element equal to 1. Then $\mathbf{A}_i$ is a vector of w_i elements, each one an attrition fraction from the appropriate state.

III. EXAMPLES

(a) The LOS Model

Let the "length of service" that a person has completed with an organization be denoted LOS, and let the FC of a person denote his LOS. If the LOS is measured in the same time units as the planning periods then each Q_i matrix is a zero matrix. In each planning period a person's LOS must increase by one unit, so that Q has the structure

$$Q = \begin{bmatrix} 0 & P_1 & & & \\ & 0 & P_2 & & \\ & & \ddots & \ddots & \\ & & & & P_{n-1} \\ & & & & 0 \end{bmatrix}.$$

By appropriate choice of the second characteristic P_i often has special structure too. Consider a hierarchical system where the "rank" or "grade" of an individual is used as his second characteristic. Then the structure of each P_i depends on the organization's promotion scheme. Assume

that in one period a person either stays in the same grade or is promoted one grade. No demotions occur. Then P_i would have the structure

$$P_i=\begin{bmatrix} x & x & & & \\ & x & x & & \\ & & \ddots & \ddots & \\ & & & x & \\ & & & & x \end{bmatrix},$$

where x represents a non-zero element.

(b) The (Grade, LOS) Model

If, as in (a) no demotions occur, and only single promotions can occur in a time period, then if i indexes the grade, and j the length of service, then Q has the structure shown in (1). Each submatrix has special structure also. Let

q_{ij}=probability a person in state (i, j) at end of one period will be in state $(i, j+1)$ at the end of the next period,

p_{ij}=probability a person in state (i, j) at the end of one period will be in state $(i+1, j+1)$ at the end of the next period.

The transition matrix Q_i has non-zero elements only immediately above the main diagonal:

$$Q_i=\begin{bmatrix} 0 & q_{i,\,l(i)} & & & \\ & 0 & q_{i,\,l(i)+1} & & \\ & & 0 & & \\ & & & q_{i,\,u(i)-1} & \\ & & & 0 & \end{bmatrix}.$$

The transition matrix P_i has non-zero elements only on a single diagonal. If $l(i+1)\geq l(i)+1$ a nd $u(i+1)\geq u(i)+1$, then P_i has the form shown below, where:

(1) the top max $\{0, l(i+1)-(l(i)+1)\}$ rows are zeros,

(2) the last max $\{0, u(i+1)-(u(i)+1)\}$ columns are zeros.

$$P_i=\begin{bmatrix} 0 & & & & & & \\ 0 & & & & & & \\ p_{i,\,l(i+1)-1} & & & & & & \\ 0 & p_{i,\,l(i+1)} & & & & & \\ & 0 & 0 & p_{i,\,u(i)} & 0 & \cdots & 0 \end{bmatrix}.$$

If $l(i+1)\leq l(i)$, the first $l(i)+1-l(i+1)$ columns of P_i are zeros. If $u(i+1)\leq u(i)$, the bottom $u(i)+1-u(i+1)$ rows of P_i are zeros. Under any circumstances P_i has only one non-zero diagonal, and we call such a matrix a diagonal matrix.

(c) The (Grade, TIG) Model

In certain applications a person's "time in grade," denoted TIG, is more important than his time in the system (LOS). If again we allow no demotions and only single promotions per period and if the FC indexes the grade and the SC the TIG for the appropriate grade, the Q has the same structure (1). Each Q_i has the same structure as in (b), but now $l(i)=1$ for each grade i. However, each matrix P_i has a single column of non-zeros,

$$P_i=\begin{bmatrix} p_{i1} & 0 & 0 \\ p_{i,u(i)} & 0 & 0 \end{bmatrix},$$

since promotions to the next highest grade lead always to a TIG of 1. Here p_{ij} is the fraction those in grade i, with time in grade i equal to j, who are promoted to grade $i+1$.

If demotions are not allowed and only single promotions can occur per period, then grade is a characteristic which can be used as the FC. A larger number of possibilities occur for the SC. In addition to those above some useful ones are (i) skill category, (ii) physical location in a multi-location organization, and (iii) educational level. Note that educational level could also be used as the FC.

IV. SOME PROBABILISTIC PROPERTIES

Let T_i be the set of states associated with FC i; thus

$$T_i=\{(i, j)|j\in J(i)\}, \qquad i=1, 2, \ldots, n.$$

Also let $w_i=u(i)-l(i)+1$, the number of states in T_i (and $J(i)$). Finally let T_0 be the single state "out of the system."

In this section we develop the probabilistic properties of:

(1) any set of states T_i,

(2) any union of consecutively indexed sets T_i, i.e.

$$\bigcup_{i=k}^{m} T_i,$$

(3) the union of all transient states, which we call T.

One of the purposes of this development is to show that the stochastic properties of Q, typically a large matrix, are readily calculated in terms of the smaller matrices Q_i and P_i, and as seen in Section III these often have extremely simple structure which leads to simple computation.

The format of this section follows closely that of Chapter 3 of Kemeny and Snell [4]. The notation (K&S, 3.__.__) indicates that a result follows from Theorem 3.__.__ in Kemeny and Snell, albeit usually not directly.

(a) First-Order Properties

Recall that we assume system matrix Q has a fundamental matrix $N=(I-Q)^{-1}$, and each element of N is the expected number of visits to the column state starting from the row state (K&S, 3.2.4). Since Q has the structure shown in (1),

$$N=\begin{bmatrix} N_1 & N_1P_1N_2 & N_1P_1N_2P_2N_3 & \cdots & \prod_{i=1}^{n-1}(N_iP_i)N_n \\ & N_2 & N_2P_2N_3 & & \prod_{i=2}^{n-1}(N_iP_i)N_n \\ & & N_3 & & \prod_{i=3}^{n-1}(N_iP_i)N_n \\ & & & & \\ & & & & N_n \end{bmatrix}, \tag{2}$$

where

$$N_i=(I-Q_i)^{-1},\ i=1,\ \ldots,\ n,$$

the fundamental matrix for FC i. Note that the large matrix N is completely determined by the matrices N_i and P_i. Thus the only matrix inversions required are those of $(I-Q_i)$, $i=1,\ \ldots,\ n$. This is of considerable computational significance because, as previously mentioned, Q is usually a large matrix.

Each matrix N_i has a probabilistic interpretation. We pursue this interpretation and show that these matrices can be used to determine other probabilistic properties of interest.

In this section we make numerous definitions and denote the k^{th} one by Dk.

Let us consider first the properties associated with a single set of states T_i and define:

D1. $v_i(j,\ m)$=expected number of visits to state $(i,\ m)$ given that FC i is entered in state $(i,\ j)$,
V_i =a $w_i\times w_i$ matrix having $v_i(j,\ m)$ as the element in row $j-l(i)+1$ and column $m-l(i)+1$.

From (2) the element of N_i in row $j-l(i)+1$ and column $m-l(i)+1$ equals the expected number of visits to state $(i,\ m)$ given that FC i is entered in state $(i,\ j)$. (K&S, 3.2.4). So, from definition D1, we have,

$$V_i=N_i. \tag{3}$$

Note that the rows and columns of N_i and V_i correspond to states in T_i in the same order as the rows and columns of Q_i.

Now define:

D2. $\tau_i(j)$=expected time in *FC* i given that *FC* i is entered in state $(i,\ j)$,

τ_i =$[\tau_i(l(i)),\ \ldots,\ \tau_i(u(i))]$, a $w_i\times 1$ vector.

The expected time spent in FC i equals the sum of the expected number of visits to the various states in FC i. From (3) and D2,

$$\tau_i(j)=\text{component } (j-l(i)+1) \text{ of } N_i\bar{1},$$

and

$$\tau_i=N_i\bar{1}, \text{ a } w_i\times 1 \text{ vector,} \tag{4}$$

where $\bar{1}$ is a vector with all components equal to one.

We next turn our attention to where the process goes when it leaves FC i. The process upon leaving T_i must enter either T_{i+1} (if $i<n$) or T_0. Next define

D3. $b_i(j, m)$=probability of entering *FC* $i+1$ in state $(i+1, m)$ given that *FC* i is entered in state (i, j)

B_i =a $w_i\times w_{i+1}$ matrix having $b_i(j, m)$ as the element in row $j-l(i)+1$, and column $m-l(i+1)+1$,

D4. $b_i(j)$ =probability of ever entering T_{i+1} given that *FC* i is entered in state (i, j),

b_i =$[b_i(l(i)), \ldots, b_i(u(i))]$, a $w_i\times 1$ vector,

D5. $b_{i0}(j)$ =probability of never entering T_{i+1} given that *FC* i is entered in state (i, j),

b_{i0} =$[b_{i0}(l(i)), \ldots, b_{i0}(u(i))]$, a $w_i\times 1$ vector.

From these definitions it follows that

$$B_i=N_iP_i, \text{ a } w_i\times w_i \text{ matrix (K\&S, 3.5.4),} \tag{5}$$

$$b_i=B_i\bar{1}, \text{ a } w_i\times 1 \text{ vector,}$$

and

$$b_{i0}=\bar{1}-b_i$$

$$=N_iA_i, \text{ a } w_i\times 1 \text{ vector.}$$

The matrix B_i is particularly useful in manpower policy analyses. For example let f_i be a $1\times w_i$ vector of the number of people entering T_i. Then f_iB_i is a $1\times w_{i+1}$ vector of the number of these people who will eventually enter T_{i+1}. (K&S, 3.3.6). Thus B_i can be used to reveal the promotion structure in either the (Grade, LOS) or (Grade, TIG) models.

Next we consider the first-order properties related to FC's i and k where $i\leq k$. Define:

D6. $b((i, j), (k, m))$=probability of entering FC k in state (k, m) given that FC i is entered in state (i, j),

B_{ik} =a $w_i\times w_k$ matrix having $b((i, j), (k, m))$ as the element in row $j-l(i)+1$ and column $m-l(k)+1$.

From definitions D3 and D6 and a simple conditioning argument we have,

$$b((i, j), (i+2, m))=\sum_{r=l(i+1)}^{u(i+1)} b_i(j, r)b_{i+1}(r, m),$$

and

$$B_{i,\,i+2}=B_iB_{i+1}.$$

Notice from D6 that B_{ii} is an identity matrix and from D3 that $B_{i,\,i+1}=B_i$. More generally it can be shown that for $i\leq k$,

$$B_{ik}=\prod_{r=i}^{k-1} B_r, \text{ a } w_i\times w_k \text{ matrix.}$$

Define:

D7. $v((i, j), (k, m))$=expected number of visits to state (k, m) given that FC i is entered in state (i, j),

V_{ik} =a $w_i \times w_k$ matrix having $v((i, j), (k, m))$ as the element in row $j-l(i)+1$ and column $m-l(k)+1$,

D8. $b_{ik}(j)$ =probability of ever entering FC k, given that FC i is entered in state (i, j),

b_{ik} =$[b_{ik}(l(i)), \ldots, b_{ik}(u(i))]$, a $w_i \times 1$ vector.

Considering each row of B_{ik} as the part of an initial probability vector that applies to T_k, we then have,

(6) $$V_{ik}=B_{ik}N_k, \text{ a } w_i \times w_k \text{ matrix (K\&S,3.5.4)},$$

and,

$$b_{ik}=B_{ik}\bar{1}, \text{ a } w_i \times 1 \text{ vector.}$$

Define:

D9. $\tau_{ik}(j)$=expected time in FC k given that FC i was entered in state (i, j),

τ_{ik} =$[\tau_{ik}(l(i)), \ldots, \tau_{ik}(u(i))]$, a $w_i \times 1$ vector.

The expected time in an FC is the sum of the expected number of visits to states in that FC, so

$$\tau_{ik}=V_{ik}\bar{1}, \text{ a } w_i \times 1 \text{ vector.}$$

This completes our study of the first-order properties related to the various FC's of the system. The foregoing definitions by no means exhaust the first-order properties of the two-characteristic model that might conceivably be of interest. It is felt, however, that these properties will often be of practical interest and that other first-order properties may be readily derived from those given above.

(b) Two Special Cases

The elements of the fundamental matrix for FC i, N_i have a somewhat different interpretation when the states in FC i have what we call the "0–1 visiting property." We say that a state has the 0–1 visiting property if the state can be visited no more than one time. Important examples of two-characteristic models in which all transient states have the 0–1 visiting property are the models in which the FC or SC is length of service or where the SC is time in grade.

If each state in T_i has the 0–1 visiting property, then the expected number of visits to a state in T_i is equal to the probability of visiting the state. The element of N_i in row $j-l(i)+1$ and column $m-l(i)+1$ may then be interpreted as the probability of visiting state (i, m) given that FC i is entered in state (i, j).

Another property of interest is the "no return property." We say that a set of states has the no return property if it is impossible to ever make a transition into the state after a transition has been made out of the state. The 0–1 visiting property implies the no return property, but they are not equivalent. For example, in modeling manpower flows in the U.S. Civil Service one might use "GS grade" as the FC and "pay step" as the SC. Each state is then a couple (grade, pay step). A

person can stay in the same pay step for more than one period, so if there are no demotions then each state would have the no return property but not the 0–1 visiting property.

If the states in T_i have the no return property then it is possible to order the states in T_i so that Q_i is upper triangular. When Q_i is upper triangular so is $I-Q_i$ and the computation of the inverse of $I-Q_i$, i.e. the fundamental matrix for FC i, N_i, is considerably easier than in the general case.

If the states in T_i have the 0–1 visiting property, then not only is N_i upper triangular but also the elements of N_i on the main diagonal are all ones.

(c) Second Moment Properties

The format in this section follows closely that of Section (a), but here we are concerned with various second moment properties of the two-characteristic model.

Define:

D10. $v_{2,i}(j, m)$=variance of the number of visits to state (i, m) given that *FC* i is entered in state (i, j),

$V_{2,i}$ =a $w_i \times w_i$ matrix having $v_{2,i}(j, m)$ as the element in row $j-l(i)+1$ and column $m-l(i)+1$.

Following (K&S, 3.3.3),

$$V_{2,i}=N_i(2(N_i)_{dg}-I)-(N_i)_{sq.}$$

where for any square matrix A, A_{dg} and A_{sq} both have the same dimensions as A; A_{dg} is defined when A is square and is formed by setting all elements in A not on the main diagonal to zero; A_{sq} is formed by squaring all the elements in A.

Define:

D11. $\tau_{2,i}(j)$=variance of the time spent in *FC* i given that *FC* i is entered in state (i, j)

$\tau_{2,i}$ =$[\tau_{2,i}(l(i)), \ldots, \tau_{2,i}(u(i))]$, a $w_i \times 1$ vector.

Following (K&S, 3.3.5),

$$\tau_{2,i}=(2N_i-I)\tau_i-(\tau_i)_{sq}.$$

Define:

D12. $v_2((i, j), (k, m))$=variance of the number of visits to state (k, m) given that *FC* i i entered in state (i,j),

$V_2(i, k)$ =a $w_i \times w_k$ matrix having $v_2((i, j), (k, m))$ as the element in row $j-l(i)+1$ and column $m-l(k)+1$,

Following (K&S, 3.3.6),

$$V_2(i, k)=V_{ik}(2(N_k)_{dg}-I)-(V_{ik})_{sq}.$$

Define:

D13. $\tau_2((i, j), k)$=variance of time spent in *FC* k given that *FC* i is entered in state (i, j),

$\tau_2(i, k)$ $=[\tau_2((i, l(i)), k), \ldots, \tau_2((i, u(i)), k)]$, a $w_i \times 1$ vector.

Following (K&S 3.3.6)),

$$\tau_2(i, k)=B_{ik}(2N_k-I)\tau_k-(\tau_{ik})_{sq}.$$

If each state in T_i has the 0–1 visiting property, then the diagonal elements of N_i are equal to 1, and,

$$(N_i)_{dg}=I,$$

$$V_{2,i}=N_i-(N_i)_{sq},$$

$$V_2(i, k)=V_{ik}-(V_{ik})_{sq}.$$

(d) Matrices of t-Step Transition Probabilities

In this section we consider the probability of being in state (k, m) t steps after being in state (i, j). The matrices of these probabilities are called the t-step transition matrices. They are used in section V to represent the stock vectors as a sum of steady-state and transient components.

Define:

D14. $m(t:(i, j), (k, m))$=probability of being in state (k, m) t steps after being in state (i, j), $t=0, 1, 2, \ldots$

$M_{ik}(t)$ =a $w_i \times w_k$ matrix having $m(t; (i, j), (k, m))$ as the element in row $j-l(i)+1$ and column $m-l(k)+1$.

The rows of $M_{ik}(t)$ are associated with states in T_i; the columns of $M_{ik}(t)$ are associated with states in T_k.

We have immediately that

$$M_{ii}(0)=I.$$

From our assumptions on the structure of Q we have,

$$M_{ik}(t)=0 \text{ if } i>k,$$

$$M_{ik}(t)=0 \text{ if } t<k-i.$$

If the process is to be in state (k, m) exactly t steps after being in state (i, j), then it must be in some state with FC k or $k-1$ exactly $t-1$ steps after being in state (i, j). Conditioning on this fact leads to the recursive equation,

(7) $$M_{ik}(t)=M_{ik}(t-1)Q_k+M_{i,k-1}(t-1)P_{k-1}, \qquad t=1, 2, \ldots$$

For any i and k the sum over t of the probability matrices $M_{ik}(t)$ gives the matrix of the expected number of visits to states with FC k starting from states with FC i. So we have,

$$\sum_{t=0}^{\infty} M_{ik}(t)=V_{ik}, \qquad i\leq k,$$

(8)

$$=0, \qquad \text{otherwise.}$$

Recall that the Q_i matrices are transient, so V_{ik} is a matrix of finite elements. This implies that,

$$\lim_{t\to\infty} M_{ik}(t)=0. \tag{9}$$

From (7) it can be shown by an inductive argument that

$$M_{ik}(t)=\sum_{r=0}^{t-1} M_{i,k-1}(t-1-r)P_{k-1}Q_k{}^r. \tag{10}$$

The t-step transition matrices provide a rather comprehensive picture of how people move through a two-characteristic system.

(e) Conditioning on Promotion When the FC is Grade

In manpower planning one is often interested in conditional probabilities, e.g. the probability of attaining grade k given that grade i is attained. The stochastic properties of the transient matrix Q under conditioning on promotion when the FC is grade are briefly developed in this section.

Define:

D15. $(i, j; t)$=the event "in state (i, j) at time t"
T_k^* =the event "a transition is made into T_k before leaving the system."

Conditioning on the event T_k^* is the same as conditioning on promotion to grade k.

Define:

D16. $q_i(i, m)=Pr[(i, m; t+1)|(i, j; t)]$
$q_i^*(j, m)=Pr[(i, m; t+1)|(i, j; t), T_{i+1}^*]$

Provided that $Pr[T_{i+1}^*|(i, j; t)]\neq 0$, we have by conditional arguments,

$$q_i^*(j, m)=q_i(j, m)\times\frac{b_i(m)}{b_i(j)}. \tag{11}$$

Define:

D17. C_i=a $w_i\times w_i$ matrix having the elements of b_i (see D4) on its main diagonal and zeros elsewhere.

We assume that promotion to grade $i+1$ is possible from every state in T_i. Under this assumption $C_i{}^{-1}$ exists. If promotion to grade $i+1$ is impossible from some state (i, j) then we must avoid conditioning on an impossible event. This is readily accomplished by temporarily treating state (i, j) as part of T_0 (out of the system) and redefining $J(i)$, Q_i, P_i and A_i accordingly.

Define:

D18. Q_i^*=a $w_i \times w_i$ matrix having $q_i^*(j, m)$ as the element in row $j-l(i)+1$ and column $m-l(i)+1$.

Then from (11) and D17,

$$Q_i^*=C_i^{-1}Q_iC_i.$$

The matrix Q_i^* is the matrix of within grade one-step transition probabilities conditioned on the attainment of grade $i+1$.

Define:

D19. $p_i(j, m)=Pr[(i+1, m; t+1)|(i, j; t)]$

$p_i^*(j, m)=Pr[(i+1, m; t+1)|(i, j; t), T_{i+1}^*]$

P_i^* =a $w_i \times w_{i+1}$ matrix having $p_i^*(j, m)$ as the element in row $j-l(i)+1$ and column $m-l(i+1)+1$.

We then have,

$$p_i^*(j, m)=p_i(j, m)\times\frac{1}{b_i(j)}.$$

Thus from D17 and D19

$$P_i^*=C_i^{-1}P_i.$$

The matrix P_i^* is the matrix of one-step promotion probabilities conditioned on the attainment of grade $i+1$.

Because $(Q_i^*)^r=C_i^{-1}Q_i^rC_i$,

the fundamental matrix for grade i is, when we condition on promotion to grade $i+1$,

$$N_i^*=(I-Q_i^*)^{-1}$$

$$=\sum_{r=0}^{\infty}(Q_i^*)^r$$

$$=\sum_{r=0}^{\infty}C_i^{-1}Q_i^rC_i$$

$$=C_i^{-1}N_iC_i.$$

Define:

D20. $v_i^*(j, m)$=expected number of visits to state (i, m) given that grade i is entered in state (i, j) and grade $i+1$ is attained.

V_i^* =a $w_i \times w_i$ matrix having $v_i^*(j, m)$ as the element in row $j-l(i)+1$ and column $m-l(i)+1$

D21. $b_i^*(j, m)$=probability of entering grade $(i+1)$ in state $(i+1, m)$ given that grade i is entered in state (i, j) and grade $i+1$ is attained.

B_i^* =a $w_i \times w_{i+1}$ matrix having $b_i^*(j, m)$ as the element in row $j-l(i)+1$ and column $m-l(i+1)+1$.

Then one may show that,

$$V_i^*=N_i^*,$$

and

$$B_i^*=N_i^*P_i^*$$

$$=C_i^{-1}B_i.$$

Note that B_i^* is simply B_i with its rows normalized, but Q_i^* is *not* simply a row normalized form of Q_i.

As with the matrices B_i, products of matrices B_i^* with successive indices are well defined: their meaning is that of a matrix B_{ik} as defined in D6 with conditioning on attainment of grade k.

The conditioned and unconditioned matrices may be used together. For example, the elements of $B_i^* B_{i+1}$ give the probabilities of entering grade $i+2$ in the column state conditioned on starting from the row state in T_i and attaining grade $i+1$.

V. EQUATIONS OF STOCKS AND FLOWS

We begin by defining the terms "stocks" and "flows," and then discuss why stocks and flows are important in manpower planning models. Next the relations between stocks and flows in a two-characteristic model are developed. Finally, we show how the stocks can be represented as the sum of a "steady-state" component and a "transient" component.

(a) Definitions and Background

A period is the interval of time from immediately after an integer value of the time parameter t up to and including the next integer value of t. A period is identified by the value of the time parameter at the end of the period. Thus,

$$\text{period } t_1 = \{t\colon t_1 - 1 < t \leq t_1\}$$

$$\text{where } t_1 \text{ is an integer.}$$

The number of people in a state at the end of a period is referred to as the "stock" in that state. Thus, stocks are counted only at integer values of the time parameter t.

The number of people who change their status in the system from one state to another during any period is referred to as a "flow." Flows occur during a period, but we do not specify the exact time at which they occur.

Stocks and flows are of primary importance in most manpower planning models. The most obvious reason for this is that costs are closely related to stocks and flows, e.g., total payroll depends on stocks; transportation costs or retraining costs depend on flows. Recruiting policy and promotion policy depend in the short term on present stocks and in the long term on how we model future stocks and flows. Determining the feasibility of a retirement plan and evaluating the effects of a change in billet structure are other instances in which the planner needs to be able to model stocks and flows in a manpower system.

We now define the variables that are used to model the stocks and flows in the two-characteristic model. Recall that T_i is the set of states associated with *FC* i, w_i is the number of states in T_i, and for convenience of notation we assume the second characteristic takes on successive integer values for *FC* i.

In a Markov model the stocks and flows are in general random variables. In this section we deal only with the expected values of stocks and flows. Such a model is called a "fractional flow

model" because the transition probabilities of the Markov model are in effect treated as fractions which direct flows through the system in a deterministic manner.

Let,

$s_{ij}(t)$=expected stocks in state (i, j) at time t,

$s_i(t)=(s_{i,l(i)}(t), \ldots, s_{i,u(i)}(t))$, a $1\times w_i$ vector of expected stocks in T_i,

$s(t)=(s_1(t), \ldots, s_n(t))$, a $1\times\sum_{i=1}^{n} w_i$ vector of expected stocks in the system.

By our basic assumption, flows into any state in T_i must come from a state in either T_i or T_{i-1}. We also make provision in our model for "external flows." The source of such flows is unspecified. However, we may consider external flows as consisting of people hired into the system The external flows may be deterministic or random, but we deal only with their expected values.

Let,

$d_{ij}(t)$=expected flow from states in T_i to state (i, j) during period t, a scalar;
$d_i(t)=(d_{i,l(i)}(t), \ldots, d_{i,u(i)}(t))$, a $1\times w_i$ vector;
$f_{ij}(t)$=expected external flow into state (i, j) during period t, a scalar;
$f_i(t)=(f_{i,l(i)}(t), \ldots, f_{i,u(i)}(t))$, a $1\times w_i$ vector;
$g_{ij}(t)$=expected flow from states in T_{i-1} to state (i, j) during period t, a scalar;
$g_i(t)=(g_{i,l(i)}(t), \ldots, g_{i,u(i)}(t))$, a $1\times w_i$ vector.

When $i=1$, $g_{ij}(t)$ is defined to be zero.

The relation between the flow vectors and the stock vector in grade i is depicted in **Figure 1**, where "T_i; t" denotes the states with FC i at time t.

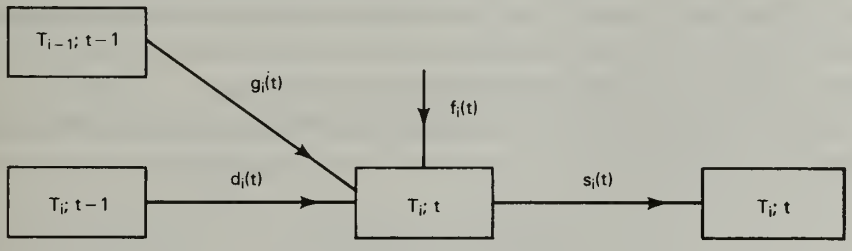


FIGURE 1. Stocks and Flows with FC i in Period t.

(b) Basic Stock Equation

Clearly, from our assumptions,

$$s_i(t)=d_i(t)+f_i(t)+g_i(t).$$

(See Figure 1.)

It will be convenient to define,

$s_0(t)=0$, a vector of zeros,

$P_0=0$, a matrix of zeros.

Using conditional expectation we then have

$$d_i(t)=s_i(t-1)Q_i, \qquad i=1, \ldots, n,$$

$$g_i(t)=s_{i-1}(t-1)P_{i-1}, \qquad i=1, \ldots, n.$$

The *basic stock equation* is then,

$$s_i(t)=s_i(t-1)Q_i+f_i(t)+s_{i-1}(t-1)P_{i-1}, \qquad i=1, \ldots, n. \tag{12}$$

The basic stock equation for FC i can be written in terms of the expected or actual stocks with FC i in previous periods. By recursively applying the basic stock equation for $s_i(t)$, $s_i(t-1)$, ..., $s_i(1)$ one obtains

$$s_i(t)=s_i(0)Q_i^t+\sum_{r=0}^{t-1} f_i(t-r)Q_i^r+\sum_{r=0}^{t-1} s_{i-1}(t-r-1)P_{i-1}Q_i^r,$$

$$t=0, 1, 2, \ldots, \tag{13}$$

$$i=1, \ldots, n,$$

which we will refer to as the *cumulative stock equation.*

Equations (12) and (13) are used frequently in the remainder of this paper. Some manpower models used in the U.S. military for short-range forecasting consist principally of an equation similar to (12).

(c) Transient Properties of the Stocks

In this section we develop a method for expressing the stock vector as a sum of a "steady-state" component and a "transient" component. This method helps one to understand how the stock vectors change in going from any present stock vector to future stock vectors. This method also helps one interpret the character of the limiting stock vector.

We do not want to restrict ourselves to cases in which the stock vector converges (as t increases) to a finite vector. We say that the vector function $\tilde{s}_i(t)$ is a steady-state component of the stock vector $s_i(t)$ if,

$$\lim_{t\to\infty} (s_i(t)-\tilde{s}_i(t))=0.$$

For any sequence of stock vectors $<s_i(t)>$ there is more than one choice of the steady-state component. In applications one would prefer a steady-state component having a relatively simple mathematical form. We show that in some cases a judicious choice of $\tilde{s}_i(0)$ makes this possible. The following theorem shows the properties of a class of steady-state components which are quite useful.

THEOREM: For any collection of $1\times w_i$ vectors $\tilde{s}_i(0)$, $i=1, \ldots, n$, let the vector functions $\tilde{s}_i(t)$ satisfy

$$\tilde{s}_i(t)=\tilde{s}_i(t-1)Q_i+f_i(t)+\tilde{s}_{i-1}(t-1)P_{i-1}, \qquad t=1, 2, \ldots,$$

$$i=1, \ldots, n,$$

i.e. the vector functions $\tilde{s}_t(t)$ satisfy the basic stock equation (12). Then

(i) the actual stocks at time t are

$$s_t(t)=\tilde{s}_t(t)+\sum_{k=1}^{i}(s_k(0)-\tilde{s}_k(0))M_{kt}(t),$$

(ii)
$$\sum_{t=0}^{\infty}(s_t(t)-\tilde{s}_t(t))=\sum_{k=1}^{i}(s_k(0)-\tilde{s}_k(0))B_{kt}N_t,$$

a $1\times w_t$ vector having finite components,

(iii) $\tilde{s}_t(t)$ is a steady-state component of the stock vector $s_t(t)$, i.e.

$$\lim_{t\to\infty}(s_t(t)-\tilde{s}_t(t))=0.$$

Before proving the theorem we explain why one might be interested in such a theorem. Part (iii) of the theorem says that $\tilde{s}_t(t)$ is a steady-state component of the stock vector $s_t(t)$, and part (i) shows how the stock vector $s_t(t)$ can be expressed as the sum of a steady-state component and a transient component. Part (ii) of the theorem says that the total over all periods of the difference between the stock vector and its steady-state component is a readily calculated finite vector.

Such information can be useful when long-range planning has been done using an "equilibrium model." As an example consider an organization which intends to change from its present size of 250,000 to a size of 200,000. The manpower planner may use an equilibrium model to develop policies that are in some sense optimal, and these policies will maintain the organization at 200,000 people once it has been reduced to this size. So the equilibrium model tells the planner what to do once the size of the organization reaches the desired equilibrium level but it doesn't tell him how to change the organization from the present level (250,000) to the desired equilibrium level (200,000). This problem of finding an optimal transition policy to go from present stock levels to a future equilibrium stock distribution is a very difficult one (see [1] Chapter 4). One method for making the transition is to immediately implement the hiring, promotion and attrition policies that have been derived from the equilibrium model. Because of the transient nature of the system these policies will eventually bring the stocks in the system to their equilibrium levels.

In the theorem the vector functions $\tilde{s}_t(t)$ play the role of what the stocks would be at time t if the system were in equilibrium. The stock vectors $s_t(t)$ indicate what the stocks will be at time t if we start with the present stocks $s_t(0)$ and implement the policies of the equilibrium model (which are reflected in the external flows, $f_t(t)$, and the transition matrices Q_t, P_t and A_t). From part (i) of the theorem we may readily calculate the difference between actual stocks and equilibrium stocks in any grade and any period. If there is a penalty associated with having more people than the equilibrium stocks in the system, then part (ii) of the theorem may be used to calculate the total penalty. Part (iii) of the theorem assures the planner that the difference between the actual and equilibrium stocks does converge to a zero vector as the time parameter t increases.

The proof of the theorem follows.

PROOF: By hypothesis the vector functions $\tilde{s}_t(t)$ satisfy the basic stock equation (12), so they must also satisfy the cumulative stock equation (13):

$$\tilde{s}_i(t)=\tilde{s}_i(0)Q_i{}^t+\sum_{r=0}^{t-1}f_i(t-r)Q_i{}^r+\sum_{r=0}^{t-1}\tilde{s}_{i-1}(t-r-1)P_{i-1}Q_i{}^r.$$

Of course the stock vectors $s_i(t)$ also satisfy the cumulative stock equation (13), so we have,

$$s_i(t)-\tilde{s}_i(t)=(s_i(0)-\tilde{s}_i(0))Q_i{}^t+\sum_{r=0}^{t-1}(s_{i-1}(t-r-1)-\tilde{s}_{i-1}(t-r-1))P_{i-1}Q_i{}^r.$$

When $i=1$ this implies,

$$\begin{aligned}s_1(t)&=\tilde{s}_1(t)+(s_1(0)-\tilde{s}_1(0))Q_1{}^t\\&=\tilde{s}_1(t)+\sum_{k=1}^{1}(s_k(0)-\tilde{s}_k(0))M_{k,1}(t),\end{aligned}$$

so we have shown that part (i) of the theorem is true when $i=1$. Suppose part (i) of the theorem is true for grade $i-1$, i.e.

$$s_{i-1}(t)=\tilde{s}_{i-1}(t)+\sum_{k=1}^{i-1}(\tilde{s}_k(0)-s_k(0))M_{k,i-1}(t).$$

Then,

$$s_{i-1}(t-r-1)-\tilde{s}_{i-1}(t-r-1)=\sum_{k=1}^{i-1}(s_k(0)-\tilde{s}_k(0))M_{k,i-1}(t-r-1),$$

and

$$\begin{aligned}s_i(t)-\tilde{s}_i(t)&=(s_i(0)-\tilde{s}_i(0))Q_i{}^t+\sum_{r=0}^{t-1}\sum_{k=1}^{i-1}(s_k(0)-\tilde{s}_k(0))M_{k,i-1}(t-r-1)P_{i-1}Q_i{}^r\\&=(s_i(0)-\tilde{s}_i(0))Q_i{}^t+\sum_{k=1}^{i-1}(s_k(0)-\tilde{s}_k(0))\sum_{r=0}^{t-1}M_{k,i-1}(t-r-1)P_{i-1}Q_i{}^r\end{aligned}$$

From equation (10) in Section IV

$$\sum_{r=0}^{t-1}M_{k,i-1}(t-r-1)P_{i-1}Q_i{}^r=M_{ki}(t),$$

so we have shown by induction that,

$$s_i(t)-\tilde{s}_i(t)=(s_i(0)-\tilde{s}_i(0))Q_i{}^t+\sum_{k=1}^{i-1}(s_k(0)-\tilde{s}_k(0))M_{ki}(t).$$

This proves part (i) of the theorem.

From part (i),

$$\begin{aligned}\sum_{t=0}^{\infty}(s_i(t)-\tilde{s}_i(t))&=\sum_{t=0}^{\infty}\sum_{k=1}^{i}(s_k(0)-\tilde{s}_k(0))M_{ki}(t)\\&=\sum_{k=1}^{i}(s_k(0)-\tilde{s}_k(0))\sum_{t=0}^{\infty}M_{ki}(t)\\&=\sum_{k=1}^{i}(s_k(0)-\tilde{s}_k(0))B_{ki}N_i,\end{aligned}$$

a $1\times w_i$ vector having finite components.

The last step above follows from equations (6) and (8) of Section IV. This proves part (ii) of the theorem.

Part (iii) follows from the fact that the sum in part (ii) is finite, and the proof of the theorem is complete.

The utility of this approach depends on our ability to find vectors $\tilde{s}_i(0)$ such that the vector functions $\tilde{s}_i(t)$ are simple and readily calculated. Some examples follow.

1. Fixed External Flows

The equilibrium models previously mentioned enjoy some popularity in military manpower planning in the United States (see for example [7]). The rationale underlying the use of such models is that one should determine the organization structure and the policies to maintain this structure which are optimal (or "least infeasible"). Among the policies derived from an equilibrium model is the hiring policy. This had the form,

$$f_i(t)=f_i \qquad t=1, 2, \ldots,$$
$$i=1, \ldots, n$$

where the vector of the number of people to be hired into the states in grade i each period, f_i, is specified from the equilibrium model.

Define,

$$\tilde{s}_1(0)=f_1N_1.$$

Then using (12) it is easy to show that

$$\tilde{s}_1(t)=f_1N_1 \text{ for all } t.$$

Thus, from the theorem

$$s_1(t)=\tilde{s}_1(t)+(s_1(0)-\tilde{s}_1(0))M_{11}(t)$$
$$=f_1N_1+(s_1(0)-f_1N_1)Q_1{}^t.$$

Now recursively define,

$$\tilde{s}_1=\tilde{s}_1(t)=f_1N_1,$$

$$\tilde{s}_i=(f_i+\tilde{s}_{i-1}P_{i-1})N_i, \qquad i=2, \ldots, n. \tag{14}$$

It is straightforward to verify that these $\tilde{s}_i$ satisfy the basic stock equation (12), so we have from the theorem, when $f_i(t)=f_i$,

$$s_i(t)=\tilde{s}_i+\sum_{k=1}^{i}(s_k(0)-\tilde{s}_k)M_{ki}(t).$$

The steady-state component can also be written,

$$\tilde{s}_i=\sum_{k=1}^{i}f_kB_{ki}N_i, \qquad i=1, \ldots, n. \tag{15}$$

Note that

$$\sum_{k=1}^{i}f_kB_{ki}$$

is a non-negative $1\times w_i$ vector, so the limiting vector of stocks in grade i must be a non-negative combination of the rows of N_i. Thus, in general, not all non-negative $1\times w_i$ vectors are possible limiting stock vectors under constant external flows.

2. Linear Growth of External Flows

In this subsection we consider the case in which the number of people hired into each state increases by the same amount each period. Such a hiring policy may not be natural over a long period of time, but it may provide a simple approximation to planned hiring policies.

Let the $1\times w_i$ vector f_i be the increase in the number hired into states with FC i each period. Then the external flow vector for FC i is,

$$f_i(t)=tf_i, \qquad t=1, 2, \ldots,$$
$$i=1, \ldots, n.$$

Let,

$$\tilde{s}_1(0)=-f_1N_1Q_1N_1.$$

Let the vector function $\tilde{s}_1(t)$ satisfy the basic stock equation (1),

$$\tilde{s}_1(t)=\tilde{s}_1(t-1)Q_1+f_1(t).$$

Using the identity $N_1Q_1+I=N_1$ one can show that

$$\tilde{s}_1(t)=tf_1N_1-f_1N_1Q_1N_1.$$

Thus from the theorem,

$$s_1(t)=tf_1N_1-f_1N_1Q_1N_1+(s_1(0)+f_1N_1Q_1N_1)Q_1^t.$$

We note that $s_1(t)$ is of the form

$$\tilde{s}_1(t)=tL_1+C_1$$
where $L_1=f_1N_1$ is a $1\times w_i$ vector,
and $C_1=-f_1N_1Q_1N_1$ is a $1\times w_i$ vector.

Consider some FC $i\epsilon\{2, \ldots, n\}$. Suppose that

$$\tilde{s}_{i-1}(t)=tL_{i-1}+C_{i-1},$$
where L_{i-1} and C_{i-1} are $1\times w_{i-1}$ vectors.

Using the identity

$$(tf_iN_i-f_iN_iQ_iN_i)Q_i+(t+1)f_i=((t+1)f_iN_i-f_iN_iQ_iN_i),$$

one may show that

$$\tilde{s}_i(t)=tf_iN_i-f_iN_iQ_iN_i+\tilde{s}_{i-1}(t-1)P_{i-1}N_i-L_{i-1}P_{i-1}N_iQ_iN_i,$$

then $\tilde{s}_i(t)$ satisfies the basic stock equation (12). Note that $\tilde{s}_i(t)$ has the form,

$$\tilde{s}_i(t)=tL_i+C_i,$$

where,

(16)
$$L_i=f_iN_i+L_{i-1}P_{i-1}N_i$$
$$=(f_i+L_{i-1}P_{i-1})N_i,$$

and,

$$C_i=-(f_i+L_{i-1}P_{i-1})N_iQ_iN_i-(L_{i-1}-C_{i-1})P_{i-1}N_i$$
$$=-((L_{i-1}-C_{i-1})P_{i-1}+f_iN_iQ_i)N_i.$$

Thus we have shown that when the external flows grow linearly, the steady-state component of the stocks also grows linearly.

By recursive substitution in (16) we have,

$$L_i=\sum_{k=1}^{i} f_kB_{ki}N_i.$$

Note that this vector gives the expected number of visits to states with FC i of $f_k=f_k(t+1)-f_k(t)$ entrants with FC k, $k=1, \ldots, i$. That is, the *growth* in the stocks with FC i each period, L_i, equals the expected number of visits to FC i of the *growth* in the external flows each period in the FC's less than or equal to i.

Both L_i and C_i have the fundamental matrix N_i as a right factor, so the steady state component of the stock vector, $s_i(t)$ must be a nonnegative combination of the rows of N_i. This same result was observed in the case of constant external flows.

In summary we have shown that by choosing

$$\tilde{s}_i(t)=tL_i+C_i$$

where

$$L_i=f_1N_1 \text{ when } i=1,$$
$$=(f_i+L_{i-1}P_{i-1})N_i, \qquad i=2, \ldots, n,$$

and

$$C_1=-f_1N_1Q_1N_1 \text{ when } i=1,$$
$$=-((L_{i-1}-C_{i-1})P_{i-1}+f_iN_iQ_i)N_i, \qquad i=2, \ldots, n,$$

then from the theorem the stock equation may be written

$$s_i(t)=\tilde{s}_i(t)+\sum_{k=1}^{i}(s_k(0)-\tilde{s}_k(0))M_{ki}(t).$$

3. Geometric Growth of External Flows

In this subsection we show that geometric growth of external flows leads (eventually) to geometric growth of the stocks. We consider the case in which the external flows into the states in grade i are proportional to a known vector f_i and grow geometrically at a rate θ_i. Thus,

$$f_i(t)=\theta_i{}^tf_i, \qquad t=1, 2, \ldots,$$
$$i=1, \ldots, n$$
$$\theta_i>0.$$

When $0<\theta_i<1$, the external flows contract rather than grow.

If θ_k is not an eigenvalue of Q_i for $k\leq i\leq n$ we may define,

$$N_i(\theta_k)=\left(I-\frac{1}{\theta_k}Q_i\right)^{-1}.$$

If the states in grade i have the 0–1 visiting property then all eigenvalues of Q_i are zero and thus $\theta_k > 0$ is never equal to an eigenvalue of Q_i in this case.

The following identity will be useful:

$$N_i(\theta_k) = \sum_{r=0}^{\infty} \left(\frac{1}{\theta_k} Q_i\right)^r$$

$$N_i(\theta_k) Q_i = \theta_k \sum_{r=0}^{\infty} \left(\frac{1}{\theta_k} Q_i\right)^{r+1}$$

$$= \theta_k(-I + N_i(\theta_k)).$$

Define,

$$\tilde{s}_1(\theta) = f_1 N_1(\theta_1).$$

Then it can be shown that if

$$\tilde{s}_1(t) = \theta_1{}^t f_1 N_1(\theta_1),$$

then $\tilde{s}_1(t)$, $t=0, 1, \ldots$, satisfies the basic stock equation, from the theorem,

$$s_1(t) = \theta_1{}^t f_1 N_1(\theta_1) + (s_1(0) - f_1 N_1(\theta_1)) M_{11}(t).$$

Note that the steady-state component of the FC 1 stock vector grows geometrically at the same rate as the external flows into FC 1.

Define,

$$B_{ki}(\theta_k) = \prod_{m=k}^{i-1} (N_m(\theta_k) P_m), \qquad 1 \leq k \leq i \leq n.$$

Then it can be shown that if

$$\tilde{s}_i(t) = \sum_{k=1}^{i} \theta_k^{t-(i-k)} f_k B_{ki}(\theta_k) N_i(\theta_k)$$

then $\tilde{s}_i(t)$, $t=0, 1, \ldots$, satisfies the basic stock equation (12). Note that in the limit the stocks with FC i grow geometrically at the rate of the largest θ_k where $k \leq i$.

Define,

$$\theta_M = \max \{\theta_k; k=1, \ldots, i\},$$

The steady-state component of the stock vector is not in general a non-negative combination of the rows of N_i (as was the case with constant external flows and linear growth of external flows). Rather the steady-state stock distribution is a non-negative combination of the rows of $N_i(\theta_M)$. The rows of $N_i(\theta_M)$ need not be non-negative combinations of the rows of N_i, so the limiting stock distributions that are possible under geometric growth of external flows need not be the same as the limiting stock distributions under constant external flows and linear growth of external flows.

BIBLIOGRAPHY

[1] Bartholomew, D. J., *Stochastic Models for Social Processes*, 2nd Edition (Wiley, 1973).

[2] Blumen, I., M. Kogan, and P. McCarthy, *The Industrial Mobility of Labor as a Probability Process* (Cornell University, 1955).

[3] Grinold, R. C. and K. T. Marshall, *Manpower Planning Models* (Elsevier-North Holland, 1977).

[4] Kemeny, J. G. and J. L. Snell, *Finite Markov Chains* (Van Nostrand, 1960).

[5] Lane, K. F. and J. E. Andrew, "A Method of Labour Turnover Analysis," Journal of the Royal Statistical Society, *A118*, 296–323 (1955).

[6] Marshall, K. T., "A Comparison of Two Personnel Prediction Models," Operations Research, *21*, (3) 810–822 (1973).

[7] RAND Corporation, "Planning in Large Personnel Systems: A Reexamination of the TOP LINE Static Planning Model," R–1274–PR (1973).

[8] Rowland, K. M. and M. G. Sovereign, "Markov-Chain Analysis of Internal Manpower Supply," Industrial Relations, *9*, (1) 88–99 (1969).

[9] U.S. Navy, "Computer Models for Manpower and Personnel Management: State of Current Technology," (NAMPS Project Report 73–2), Naval Personnel Research and Development Laboratory (1973).

JOINT PRICING AND ORDERING POLICY FOR EXPONENTIALLY DECAYING INVENTORY WITH KNOWN DEMAND

Morris A. Cohen
The Wharton School
University of Pennsylvania
Philadelphia, Pennsylvania

ABSTRACT

This paper is concerned with the problem of simultaneously setting price and production levels for an exponentially decaying product. Such products suffer a loss in utility which is proportional to the total quantity of stock on hand. A continuous review, deterministic demand model is considered. The optimal ordering decision quantity is derived and its sensitivity to changes in perishability and product price is considered. The joint ordering pricing decision is also computed and consideration of parametric changes of these decisions indicates a non-monotonic response for optimal price to changes in product decay. Issues of market entry and extensions to a model with shortages are also analyzed.

INTRODUCTION

Analysis of inventories of goods whose utility does not remain constant over time has involved a number of different concepts of deterioration. It is possible to identify problems in which all items in the inventory become obsolete at some fixed point in time (the style good problem), and problems where the product deteriorates throughout the planning horizon. The class of products subject to on-going deterioration can be broken down into those products with a maximum usable lifetime (perishable products) and those without (decaying products).

Thus, for example, spare parts for military aircraft are style goods since they become obsolete when a replacement model is introduced. Both blood and certain foods are examples of perishable products with a maximum usable lifetime. Volatile liquids such as alcohol and gasoline are products which decay and which do not have a maximum lifetime.

The decrease in utility or loss for an inventory of goods subject to deterioration is usually a function of the total amount of inventory on hand. For goods without a maximum lifetime the items in the inventory can be grouped together for the purpose of determining how much stock will decay at a given point in time. It is clear that the amount of inventory which deteriorates in the case where there is a maximum lifetime is a function of the age distribution of all items in the inventory.

This paper will be concerned with the problem of simultaneously setting price and production levels for an exponentially decaying product. The earliest attempts at modeling such problems were concerned with optimal production decisions only. Ghare and Schrader [6], assuming expo-

nential decay of the inventory in the face of constant demand, derived a revised form of the economic order quantity. Emmons [5] also considered a problem of exponential decay where the product decayed at one rate into a new product which decayed at a second rate. These models are applicable to inventories of radioactive isotopes. In the first part of this thesis, Van Zyl [16] formulated a general age independent perishable good model in which a fixed or stochastic amount of product, depending on the total inventory, deteriorates. He demonstrated that for this class of inventory models the optimal order policy is of the fixed critical number form and is thus characterized by a constant order-up-to quantity.

The analysis of price as an inventory decision variable for a nonperishable product has been undertaken by a number of authors (Whitin [17], Thomas [14], Karlin and Carr [7], Kunreuther and Richard [8], Kunreuther and Schrage [9], Adams [1] and Pekelman [13]). For the most part this literature has been concentrated on deterministic models (with some exceptions [7, 14]). The only example of an analysis of pricing policy for a perishable product is due to Eilon and Mallya [4], in which fairly strong assumptions are made on the form of the issuing sequence in force for a maximum lifetime (perishable) product model. The analysis of ordering policies for perishable products has been considered extensively in the last few years (see Cohen [2] for a recent survey).

1. NO SHORTAGE MODEL

We shall consider in this section a continuous review, deterministic demand model with exponential decay. This model will provide some intuitive insights and analytic results which will be useful in analysing the economic tradeoffs inherent in the control of perishable commodities. We begin with a no shortage assumption. In the section to follow, this assumption is relaxed and the deterministic model is extended to allow for shortages with backlogging.

Let p stand for the selling price of the product and $d(p)$ for the known demand rate when the price is p. Let $I(t)$ be the inventory position at time t and λ a positive number representing the stock decay rate. As noted previously, perishability will be of the exponential type and hence the rate at which stock decays will be proportional to the on hand inventory, $I(t)$. Demand rate $d(p)$ is assumed to be positive and to possess a negative derivative throughout its domain. In the case of continuous review it is logical to assume that depletion due to such decay and depletion due to meeting demand will occur simultaneously. Accordingly, the differential equation describing the time behavior of the system is:

$$\frac{dI(t)}{dt} = -\lambda I(t) - d(p). \tag{1}$$

It follows from the fact that this is a first order linear differential equation (as noted in Ghare and Schrader [6]) that the solution to (1) is

$$I(t) = I(0)e^{-\lambda t} - (d(p)/\lambda)[1 - e^{-\lambda t}]. \tag{2}$$

Consequently, $Z(t)$, stock loss due to decay in the time interval $[0, t]$ is the difference between the inventory position at time t which would prevail if there was no decay and the position with the decay.

$$Z(t) = I(t)[e^{\lambda t} - 1] - d(p)t + (d(p)/\lambda)[e^{\lambda t} - 1]. \tag{3}$$

Since the cost structure to be defined includes holding costs, it is clearly optimal to set $I(T)=0$, where T is the period length of each cycle. Figure 1 illustrates the time behavior of the inventory level. Demand rate, $d(p)$ is indicated by the slope of the dashed line.

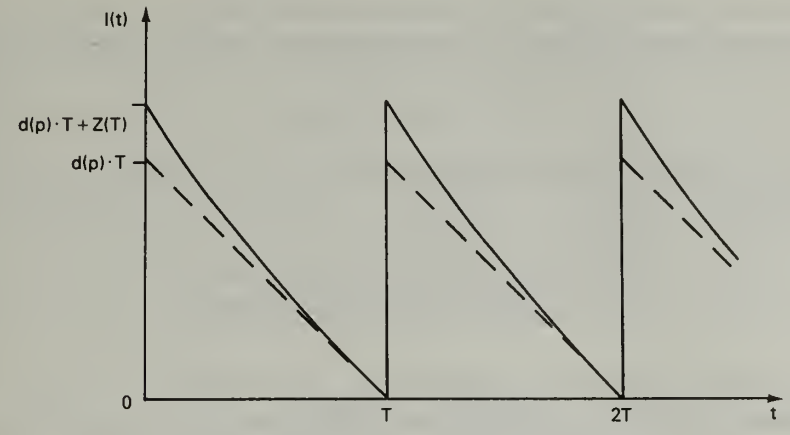


FIGURE 1. Inventory Level vs. Time (No Shortages).

Given that $I(T)=0$ and noting that the only loss to the system is due to either decay or demand, the following expression for the quantity ordered each cycle results:

$$\begin{aligned} Q_T &= Z(T)+d(p)T \\ &= -d(p)T+(d(p)/\lambda)[e^{\lambda T}-1]+d(p)T \\ &= (d(p)/\lambda)[e^{\lambda T}-1]. \end{aligned} \tag{4}$$

Noting that $I(0)=Q_T$ it also follows that

$$I(t)=\frac{d(p)}{\lambda}\,[e^{\lambda(T-t)}-1]. \tag{5}$$

Let us define unit purchase cost, order set-up cost and unit holding cost by c, K and h dollars, respectively. Cost per cycle then becomes,

$$\begin{aligned} \hat{C}(T,p) &= K+cQ_T+h\int_0^T I(t)\,dt \\ &= K+c(d(p)/\lambda)[e^{\lambda T}-1]+h(d(p)/\lambda^2)[e^{\lambda T}-1-\lambda T] \end{aligned} \tag{6}$$

for a fixed price level p. Cost per unit time, $C(T,p)$, is

$$\begin{aligned} C(T,p) &= \hat{C}(T,p)/T \\ &= K/T+([c\lambda+h]d(p)(e^{\lambda T}-1))/\lambda^2 T-hd(p)/\lambda. \end{aligned} \tag{7}$$

By holding p fixed we can consider the necessary conditions for minimizing $C(T,p)$ with respect to T.

$$\frac{\partial C(T,p)}{\partial T}=\frac{1}{T^2}\{-K+(c\lambda+h)d(p)[\lambda e^{\lambda T}T-e^{\lambda T}+1]/\lambda^2\}=0. \tag{8}$$

Which implies that

$$e^{\lambda T}(\lambda T-1)=K\lambda^2/[d(p)[c\lambda+h]]-1. \tag{9}$$

(9) can be easily solved numerically for T_p, the cost-minimizing† cycle length. An approximate solution to (9) can be obtained by using a truncated Taylor series expansion for the exponential function, i.e.,

$$e^{\lambda T}\approx 1+\lambda T+\lambda^2 T^2/2,$$

which is a valid approximation for smaller values of λT. It follows that (7), the definition of cost per unit time, becomes,

$$\begin{aligned} C(T, p) &\approx cd(p)[1+\lambda T/2]+K/T+hd(p)T/2 \\ &= cd(p)+(c\lambda+h)d(p)T/2+K/T. \end{aligned} \tag{10}$$

Thus it is clear that the Taylor Series approximation for $e^{\lambda T}$ yields an inventory model without decay but with holding cost $c\lambda+h$. The derivative of the approximate cost function is,

$$\frac{\partial C(T, p)}{\partial T}=cd(p)\lambda/2-K/T^2+hd(p)/2=0$$

and so,

$$T_p=\sqrt{2K/(d(p)[c\lambda+h])}. \tag{11}$$

For fixed selling price p, the optimal cycle length decreases as decay rate λ increases. Moreover, since demand $d(p)$ has been assumed to decrease with increasing p, it is clear that the cycle length will increase as the price increases. Thus highly perishable goods facing high demand will be replenished more often. We note as well that when $\lambda=0$ there is no perishability and (11) reduces to the standard form.

We can also consider the effects of variation in product perishability and price changes on the optimal order decision. For comparative purposes we examine the optimal order rate. From (4) and (11),

$$\begin{aligned} Q_{T_p}/T_p &= d(p)[e^{\lambda T_p}-1]/\lambda T_p \\ &\approx d(p)[1+\lambda T_p/2] \end{aligned} \tag{12}$$

where the first term corresponds to the demand rate and the second term approximates the rate at which units spoil. The sensitivity of the order rate to changes in perishability is determined by,

$$\begin{aligned} \frac{\partial}{\partial\lambda}[Q_{T_p}/T_p] &= d(p)T_p/2+\lambda d(p)\frac{\partial T_p}{\partial\lambda}/2 \\ &= \sqrt{Kd(p)/2[c\lambda+h]}\,[(\lambda c+2h)/(2\lambda c+2h)] \\ &> 0. \end{aligned}$$

† $C(T, p)$ is in fact not convex for all values of the cost parameters. A sufficient condition for convexity is

$$e^{\lambda T}[1+(\lambda T-1)^2]+2K\lambda^2/(c\lambda+h)d(p)>4.$$

This condition will be easily satisfied with relatively high set up costs, K or larger values of λ.

Similarly, order rate will respond to a price change as,

$$\frac{\partial}{\partial p}[Q_{T_p}/T_p]=d'(p)[1+\lambda T_p/4]\leq 0.$$

We see then that optimal order rate increases with an increase in decay rate λ and decreases with increasing price when we assume price to be an external (market-controlled) parameter.

The validity of the Taylor series approximation for the exponential function decreases when values of λ closer to 1 prevail. It is important to verify that the response of optimal cycle length and order rate to changes in both price p and decay rate λ is consistent with the results derived for the approximate cost function. An example problem was considered by solving (9) directly for T_p. The corresponding optimal order rate Q_{T_p}/T_p was then computed. The results of the experiment and the associated values of the cost parameters and values of p and λ are illustrated in Table 1.

The expected reactions, i.e.,

$$\frac{\partial T_p}{\partial \lambda}\leq 0, \qquad \frac{\partial T_p}{\partial p}\geq 0,$$

$$\frac{\partial[Q_{T_p}/T_p]}{\partial \lambda}\geq 0 \quad \text{and} \quad \frac{\partial[Q_{T_p}/T_p]}{\partial p}\leq 0$$

were all observed. The computations were repeated with the same results for a number of different choices of the cost parameters.

TABLE 1. Optimal cycle length and order rate, for example with $K=\$250, c=\$1$/unit and $h=\$.5$/unit/day and the demand rate function is, $d(p)=25-.5p$ (unit/day)

p \ λ	.05	.1	.4	.6	.8
0.0	5.50 28.77	4.89 32.24	3.07 49.15	2.51 58.20	2.14 66.20
1.0	5.54 28.20	4.93 31.67	3.09 48.41	2.52 57.38	2.15 65.31
4.0	5.71 26.62	5.07 29.95	3.16 46.20	2.58 54.9	2.19 62.63
6.0	5.82 25.54	5.16 28.80	3.21 44.71	2.61 53.25	2.22 60.82
10.0	6.08 23.34	5.38 26.48	3.31 41.7	2.69 49.88	2.29 57.14
15.0	6.46 20.66	5.68 23.67	3.46 37.84	2.80 45.56	2.38 52.41
20.0	6.93 17.92	6.06 20.62	3.64 33.9	2.93 41.11	2.48 47.53

In order to consider the optimal price decision we define the profit rate as a function of cycle length and price,

$$\pi(T, p)=pd(p)-C(T, p) \tag{13}$$
$$\approx pd(p)-cd(p)-K/T-[c\lambda+h]d(p)\,T/2$$

again using the exponential function approximation. Necessary conditions for maximizing π with respect to T and p yield,

$$p_T=c(1+\lambda T_p/2)+hT_p/2-d(p)/d'(p). \tag{14}$$

Under various assumptions on the form of $d(p)$, equations (11) and (14) can be solved simultaneously by numerical methods for the optimal price and period length (p^*, T^*). For example if $d(p)=A+Bp$ for $B<0$, it follows that

$$p_T=(c\lambda+h)T_p/4+c/2-A/2B \tag{15}$$

and consequently, for a fixed period length, optimal price will increase with an increase in decay rate λ or costs c and h.

Define $\pi(T_p, p)$ to be the profit function with the optimal period length in effect. The joint price/production problem is then equivalent to

$$\max_{p\geq 0} \pi(T_p, p).$$

Using (11) and approximation (13),

$$\pi(T_p, p)=pd(p)-cd(p)\{1+(\lambda/2)\sqrt{2K/d(p)[c\lambda+h]}\}$$
$$-\sqrt{Kd(p)[c\lambda+h]/2}-(h/2)\sqrt{2Kd(p)/[c\lambda+h]}$$
$$=pd(p)-cd(p)-\sqrt{2K[c\lambda+h]d(p)}, \tag{16}$$

which, as expected is the standard expression for revenue less inventory holding costs in a system with unit holding cost $c\lambda+h$. We note that,

$$\frac{\partial\pi(T_p, p)}{\partial p}=d(p)+d'(p)[p-c-\sqrt{K[c\lambda+h]/2d(p)}]=0$$

yields

$$p=-d(p)/d'(p)+c+\sqrt{K[c\lambda+h]/2d(p)} \tag{17}$$

which can be solved by successive approximation for various choices of demand rate function d. We note as well that at $p=c$,

$$\pi(T_c, c)=-\sqrt{2K[c\lambda+h]d(c)}$$

and so

$$\pi(T_p, p)<0 \qquad \text{for } p\leq c \tag{i}$$

and also

$$\left.\frac{\partial\pi(T_p, p)}{\partial p}\right|_{p=c}=d(c)-d'(c)\sqrt{K[c\lambda+h]/2d(c)}\geq 0. \tag{ii}$$

Therefore if we assume that $d(p)$ belongs to the class of functions satisfying

(i) $$d'(p) \leq 0$$

(ii) $$\lim_{p \to \infty} d(p) = 0$$

(iii) $$\lim_{p \to \infty} pd(p) = 0,$$

then $\lim_{p \to \infty} \pi(T_p, p) = 0$.

This class is fairly general and includes the truncated linear demand and the exponential demand families. Hence the following has been established.

PROPOSITION 1: $\pi(T_p, p)$ achieves its maximum at some possibly infinite $p^* > c$ and is equal to

$$\pi(T_{p^*}, p^*) = p^* d(p^*) - cd(p^*) - \sqrt{2K[c\lambda + h]d(p^*)}.$$

Thus it is possible to solve the joint price-production problem for a fairly extensive class of demand functions. We note that maximum profit may be zero, in which case p^* approaches infinity.

The variability of the optimal solution, (p^*, T^*) to changes in decay rate λ can also be investigated numerically for various cost coefficient configurations. Analysis of the previously discussed example indicates that optimal price and order period do not behave monotonically with respect to λ. These nonintuitive results are illustrated in Table 2.

The issue of market entry for a price setting monopolist facing inventory costs and a downward sloping demand curve was considered by Kunreuther and Richard [8]. The monopolist will enter the market only when he can set price to achieve a strictly positive profit. This requirement leads to the following relationship in our problem:

$$\frac{p-c}{p} > \frac{\sqrt{2K(c\lambda + h)d(p)}}{pd(p)}$$

The fractional mark up must exceed the ratio of inventory holding costs to revenue. This will be achieved in an interval of prices contained in $[c, \infty)$ and hence the optimal price will be achieved at some finite price strictly greater than c. As λ increases the producer will adjust his optimal price

TABLE 2. Variation in optimal solution with respect to decay rate λ, for example with K=\$0, c=\$1/unit, h=\$.5/unit/day and demand function $d(p) = 25 - .5p$ (unit/day)

λ	Optimal Price P^*	Optimal Cycle Length T^*	Optimal Order Rate Q^*	Profit at Optimal
.05	26.26	5.52	13.67	234.74
.10	26.50	6.67	11.71	232.40
.15	26.47	5.94	18.99	226.29
.20	26.44	5.37	21.13	220.38
.25	26.42	4.90	23.15	214.63
.30	26.41	4.53	25.09	209.03
.40	26.39	4.21	26.93	203.56
.50	26.37	3.49	32.08	187.75
.60	26.37	3.15	35.16	177.63

to remain profitable. As illustrated in Table 2, the possibility of positive profit decreases with higher values of λ. Thus increased perishability will impose a barrier for market entry on the part of the entrepreneur since profits fall as the product becomes more perishable. It is important to note that while the optimal price and cycle length decisions do not react monotonically to increases in λ, there is a marked stability in the value of the optimal price. Thus the tradeoff between revenue and loss due to decay may lead to an unexpected pattern of pricing and ordering decisions. For the particular example of Table 2 we observe that for low values of λ the optimal reaction to increased perishability is to increase price. In the range of higher values for λ an optimal reaction to increased perishability is to decrease price. Cycle length decreases with λ, and order rate increases after an initial decrease.

We turn next to the extension of the model to the case where shortages are allowed to occur.

2. MODEL WITH SHORTAGES

The model of the previous section is now extended to allow for complete backlogging of excess demands. Figure 2 illustrates the time behavior of inventory in this case. Stock is depleted by a combination of spoilage and demand in the interval $[0, T_1]$ and is backlogged as a result of excess demand in the interval $[T_1, T]$. The loss of stock due to decay within the cycle of length T is given by:

$$Z(T_1) = -d(p)T_1 + d(p)[e^{\lambda T_1} - 1]/\lambda. \tag{18}$$

Backlogged demand within the cycle, B, is defined by:

$$B(T_1) = d(p)[T - T_1] \tag{19}$$

and order quantity, Q, is the sum of satisfied demand, backlogged demand and loss due to decay.

$$Q = Z(T_1) + d(p)T. \tag{20}$$

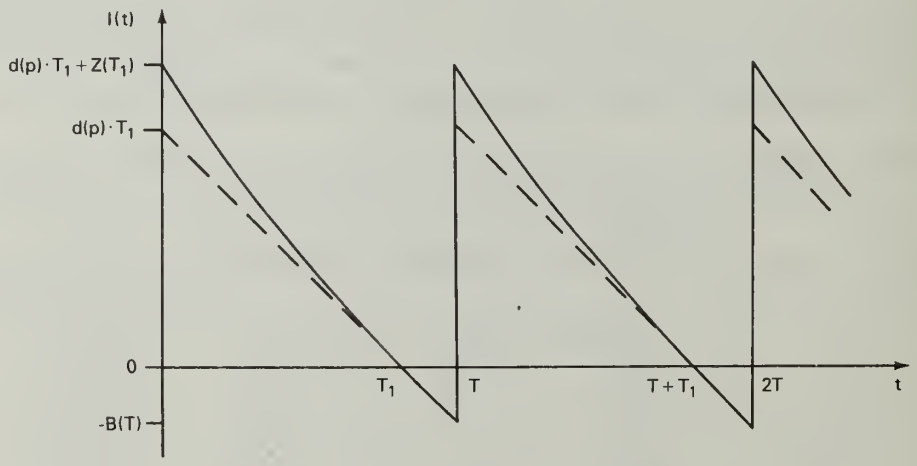


FIGURE 2. Inventory Level vs Time (Shortages).

By using the same cost structure as before with the addition of s as the unit shortage cost rate (to avoid confusion with price p) we can derive cost per cycle as follows.

$$\hat{C}(T, T_1, p)=K+cQ+h\int_0^{T_1} I(t)dt+s\int_0^{T-T_1} d(p)tdt$$
$$=K+cd(p)\{T-T_1+[e^{\lambda T_1}-1]/\lambda\}+hd(p)[e^{\lambda T_1}-1-\lambda T_1]/\lambda^2+sd(p)[T-T_1]^2/2.$$

We can also derive cost per unit time by dividing the above expression by total cycle length T:

$$(21)\qquad C(T, T_1, p)=\hat{C}(T, T_1, p)/T$$
$$=K/T+cd(p)\left\{\frac{T-T_1}{T}+\frac{e^{\lambda T_1}-1}{\lambda T}\right\}+hd(p)\frac{e^{\lambda T_1}-1-\lambda T_1}{\lambda^2 T}+sd(p)\frac{(T-T_1)^2}{2T}$$

Using the previously defined approximation for the exponential function yields the following:

$$(22)\qquad C(T, T_1, p)\approx cd(p)\left\{\frac{\lambda T_1+\lambda^2 T_1^2/2}{\lambda T}+\frac{T-T_1}{T}\right\}+K/T+hd(p)\frac{\lambda^2 T_1^2/2}{\lambda^2 T}+sd(p)\frac{(T-T_1)^2}{2T}$$

$$=cd(p)\left(\frac{T+\lambda T_1^2/2}{T}\right)+K/T+hd(p)T_1^2/2T+sd(p)(T-T_1)^2/2T.$$

Let $T_1/T=\eta$ be the fraction of the cycle in which there is no excess demand. Cost per unit time can then be expressed as a function of (T, η, p) as follows:

$$(23)\qquad C(T, \eta, p)=cd(p)+K/T+[c\lambda\eta^2+h\eta^2+s(1-\eta)^2]d(p)T/2.$$

For fixed price, p, $C(T, \eta, p)$ must be minimized with respect to T and η

$$\frac{\partial C(T, \eta, p)}{\partial T}=cd(p)\lambda\eta^2/2-K/T^2+hd(p)\eta^2/2+sd(p)(1-\eta)^2/2=0$$

which yields,

$$(24)\qquad T_{p,\eta}=\sqrt{2K/d(p)[c\lambda\eta^2+h\eta^2+s(1-\eta)^2]}.$$

We can restrict the model to exclude shortages by setting $\eta=1$, which yields the previously derived result (11).

$$\frac{\partial C(T, \eta, p)}{\partial \eta}=0$$

yields the result

$$(25)\qquad \eta=\frac{s}{c\lambda+h+s}$$

and so $\lim_{s\to\infty}\eta=1$ as expected. It is interesting to note that as λ decreases and hence as the product becomes less perishable, η increases and thus the fraction of the cycle spent backlogging demand decreases. Thus decreased perishability has the effect of raising the relative shortage cost. The analysis of optimal order rate can be carried out as in the case of the no shortage model. Similar conclusions can be derived.

In order to analyze the optimal price decision we must again consider the profit rate function.

$$\pi(T, \eta, p)=pd(p)-C(T, \eta, p)$$

Differentiating with respect to p,

$$\frac{\partial\pi(T, \eta, p)}{\partial p}=d(p)+pd'(p)-\frac{\partial C(T, \eta, p)}{\partial p}$$

$$=d(p)+d'(p)[p-c(1+\lambda\eta^2T/2)-h\eta^2T/2-s(1-\eta)^2T/2]$$

$$=0$$

yields,

$$P_{T,\eta}=c(1+\lambda\eta^2T/2)+h\eta^2T/2+s(1-\eta)^2T/2-d(p)/d'(p) \tag{26}$$

For the special case of a linear demand function,

$$P_{T,\eta}=[c\lambda\eta^2+h\eta^2+s(1-\eta)^2]T/4+c/2-A/2B.$$

Thus, again for this case, for fixed values of T and η, the optimal price decreases as λ decreases.

We can also examine the revenue function when optimal cycle length $T_{p,\eta}$ is in force. This yields the following,

$$\pi(T_{p,\eta}, p)=pd(p)-cd(p)-cd(p)\lambda\eta^2T_{p,\eta}/2-K/T_{p,\eta}-hd(p)\eta^2T_{p,\eta}/2-sd(p)(1-\eta)^2T_{p,\eta}/2$$

$$\pi(T_{p,\eta}, p)=pd(p)-cd(p)-\frac{cd(p)\lambda\eta^2}{2}\sqrt{2K/d(p)[c\lambda\eta^2+h\eta^2+s(1-\eta)^2]}-\sqrt{Kd(p)[c\lambda\eta^2+h\eta^2+s(1-\eta)^2]}/2$$

$$-\frac{hd(p)\eta^2}{2}\sqrt{\frac{2K}{d(p)[c\lambda\eta^2+h\eta^2+s(1-\eta)^2]}}-\frac{sd(p)(1-\eta)^2}{2}\sqrt{\frac{2K}{d(p)[c\lambda\eta^2+h\eta^2+(1-\eta)^2]}}$$

$$=pd(p)-cd(p)-\sqrt{2Kd(p)[c\lambda\eta^2+h\eta^2+s(1-\eta)^2]}$$

which is again revenue less inventory carrying costs. First order conditions for a maximum yield,

$$\frac{\partial\pi(T_{p,\eta}, p)}{\partial p}=d(p)+d'(p)\left[p-c-\sqrt{\frac{K[c\lambda\eta^2+h\eta^2+s(1-\eta)^2]}{2d(p)}}\right]=0$$

and thus the revenue maximizing price satisfies

$$p=-d(p)/d'(p)+c+\sqrt{K[c\lambda\eta^2+h\eta^2+s(1-\eta)^2]/2d(p)}. \tag{26}$$

which is a direct extension of (17). At $p=c$,

$$\pi(T_{c,\eta}, c)=-\sqrt{2Kd(c)[c\lambda\eta^2+h\eta^2+s(1-\eta)^2]}<0$$

and

$$\frac{\partial\pi(T_{p,\eta}, p)}{\partial p}\bigg|_{p=c}=d(c)-d'(c)\sqrt{\frac{K[c\lambda\eta^2+h\eta^2+s(1-\eta)^2]}{2d(c)}}\geq 0,$$

It also follows that when $\lim_{p\to\infty} pd(p)=0$ holds, that $\lim_{p\to\infty}\pi(T_{p,\eta}, p)=0$ as in the no shortage case. Thus the following result has been established.

PROPOSITION 2: $\pi(T_{p,\eta}, p)$ achieves its maximum in the shortage case at a possibly infinite $p^*>c$ (solving (26)) and is equal to,

$$\pi(T_{p^*,\eta}, p^*)=p^*d(p^*)-cd(p^*)-\sqrt{2d(p^*)K[c\lambda\eta^2+h\eta^2+s(1-\eta)^2]}.$$

The condition for market entry which will ensure that maximum profit is achieved at a finite price is given by

$$\frac{p-c}{p}>\sqrt{\frac{2K[C\lambda\eta^2+h\eta^2+s(1-\eta)^2]d(p)}{pd(p)}}$$

Thus we have seen that the shortage model represents a direct extension of the no shortage model. The influence of perishability, through λ, has now become more complex due to the dependence of p^*, $T_{p^*,\eta}$ and η on λ.

CONCLUSIONS

The deterministic model presented in this paper can be extended to the case of stochastic demand by making assumptions on the nature of the process generating the demand. Multiplicative and additive demand factor models were considered in a paper on pricing by Karlin and Carr [7] in which there was no perishability. Leyland [10] extended this analysis to an arbitrary demand process. The impact of perishability on a stochastic model with pricing as a decision variable was recently considered (after the preparation of this paper) by Thowsen [15], in which the existence of an optimal joint production-price policy for a stochastic demand model is demonstrated. It should be pointed out that the yet to be solved pricing problem for the maximum lifetime (perishable) product system represents a fundamental problem since the utility of the consumer with respect to aged goods must be taken into account.

This paper then is a first step in analyzing the interaction effect of perishability with optimal pricing and ordering decisions. The results indicate the importance of formulating and solving an appropriate inventory model when perishability and the opportunity for pricing occur simultaneously.

BIBLIOGRAPHY

[1] Adams, C. R., "A Monopolist's Revision of Mathematical Inventory Theory," presented at the 43rd National Meeting, Operations Research Society of America (1973).

[2] Cohen, M. A., *Inventory Control for a Perishable Product Optimal Critical Number Ordering and Applications to Blood Inventory Management*, Ph.D. dissertation (Northwestern University, Evanston, Ill., 1974).

[3] Cohen, M. A., "Analysis of Single Critical Number Ordering Policies for Perishable Inventories," Operations Research *24*, 726–741 (1976).

[4] Eilon, S. and R. V. Mallya, "Issuing and Pricing Policies for Semi-perishables," *Proceedings of 4th International Conference on Operational Research* (Wiley-Interscience, 1966).

[5] Emmons, H., "A Replenishment Model for Radioactive Nuclide Generators," Management Science, *14*, 263–274 (1968).

[6] Ghare, P. M. and G. F. Schrader, "A Model for an Exponential Decaying Inventory," The Journal of Industrial Engineering *14*, 238–243 (1963).

[7] Karlin, S. and C. B. Carr, "Prices and Optimal Inventory Policy," in *Studies in Applied Probability and Management Science*, ed. K. J. Arrow, S. Karlin and H. Scarf (Stanford University Press, Stanford, 1962).

[8] Kunreuther, H. and J. F. Richard, "Optimal Pricing and Inventory Decisions for Non-Seasonal Items," Econometrica, *39*, 173–175 (1975).

[9] Kunreuther, H. and L. Schrage," Joint Pricing and Inventory Decisions for Constant Priced Items," Management Science *7*, 732–738 (1973).

[10] Leland, H. E., "Theory of the Firm Facing Uncertain Demand," American Economic Review, *52*, 278–291 (1972).

[11] Nahmias, S., "Myopic Approximations for the Perishable Inventory Problem," Management Science, *9*, 1002–1008, (1976).

[12] Nahmias, S. and W. Pierskalla, "A Two Product Perishable/Non-Perishable Inventory Problem," SIAM Journal of Applied Mathematics, *30*, 483–500 (1976).

[13] Pekelman, D., "Simultaneous Price-Production Decisions," Operations Research, *22*, 788–794 (1974).

[14] Thomas, L. J., "Price and Production Decisions with Random Demand," Operations Research, *22*, 513–518 (1974).

[15] Thowsen, G. T., "A Dynamic, Nonstationary Inventory Problem for a Price/Quantity Setting Firm," Naval Research Logistics Quarterly, *22*, 461–476.

[16] Van Zyl, G., *Inventory Control for Perishable Commodities*. Ph.D. dissertation (University of North Carolina, 1964).

[17] Whiten, T. M., "Inventory Control and Price Theory," Management Science, *2*, 61–68 (1955).

ESTIMATION OF ORDERED PARAMETERS FROM k STOCHASTICALLY INCREASING DISTRIBUTIONS*

Hubert J. Chen

The University of Georgia
Athens, Georgia

ABSTRACT

There are given k (≥ 2) univariate cumulative distribution functions (c.d.f.'s) $G(x; \theta_i)$ indexed by a real-valued parameter θ_i, i=1, . . ., k. Assume that $G(x; \theta_i)$ is stochastically increasing in θ_i. In this paper interval estimation on the i^{th} smallest of the θ's and related topics are studied. Applications are considered for location parameter, normal variance, binomial parameter, and Poisson parameter.

1. INTRODUCTION

Suppose that we have k populations $\pi_1, \ldots, \pi_k$, with cumulative distribution functions (c.d.f.'s) $F(x; \theta_i)$ $(1 \leq i \leq k)$, where θ_i is a single parameter of $F(x; \theta_i)$. The goal is interval estimation of the i^{th} smallest of $\theta_1, \ldots, \theta_k$, denoted by $\theta_{[i]}$ which is unknown. Let $\underset{\sim}{X}_i$, $i=1, \ldots, k$, be mutually independent random samples, each of size n, from population π_i, $i=1, \ldots, k$, respectively. Let $T_i = T_n(\underset{\sim}{X}_i)$ be an estimator of θ_i with c.d.f. $G_n(t; \theta_i)$. It is assumed that the family $G_n(t; \theta)$ is stochastically increasing in θ, that is, $G_n(t; \theta') \geq G_n(t; \theta'')$ if $\theta' < \theta''$ for all t. Applications are considered when θ is a location parameter, a scale parameter (e.g., normal variance), and a parameter of a specific family of monotone likelihood ratio, e.g., a binomial parameter and a Poisson parameter. Further applications are mentioned when θ is a noncentrality parameter of a noncentral t-distribution and when θ is a correlation coefficient of a bivariate normal distribution.

In recent years several authors have considered the interval estimation of ranked parameters, such as Chen and Dudewicz [3], Dudewicz [4, 5, 6], Dudewicz and Tong [8], Saxena and Tong [15], Saxena [14], Alam, Saxena and Tong [2], and Rizvi and Saxena [13]. Alam and Saxena [1] have also considered the parameter in stochastically increasing family.

2. STOCHASTICALLY INCREASING FAMILY

In this section we give a basic method for construction of confidence intervals for $\theta_{[i]}$ $(1 \leq i \leq k)$ by using the i^{th} smallest of $T_1, \ldots, T_k$, denoted by $T_{[i]}$.

LEMMA (2.1): For any i $(1 \leq i \leq k)$, the c.d.f. of the i^{th} smallest statistic $T_{[i]}$, $H_{T_{[i]}}(t)$, is a non-increasing function of θ_l $(1 \leq l \leq k)$.

*This research was supported by the U.S. Army Research Office—Durham.

PROOF: Fix $l(1\leq l\leq k)$, for $i=1, \ldots, k$, and t is between T_* and T^* the lower and the upper limits of $T_{[i]}$, let $\underset{\sim}{\theta}=(\theta_1, \ldots, \theta_k)$,

$$H_{T_{[i]}}(t)=P_{\underset{\sim}{\theta}}(\text{At least } i \text{ of } T_1, \ldots, T_k \text{ are} \leq t)$$

$$=P_{\underset{\sim}{\theta}}(T_l\leq t \text{ and at least } i-1 \text{ of } T_1, \ldots, T_{l-1}, T_{l+1}, \ldots, T_k \text{ are} \leq t)+P_{\underset{\sim}{\theta}}(T_l>t \text{ and at least } i \text{ of}$$

$$T_1, \ldots, T_{l-1}, T_{l+1}, \ldots, T_k \text{ are } \leq t)$$

$$=G_n(t; \theta_l)\{P_{\underset{\sim}{\theta}}(\text{At least } i-1 \text{ of } T_1, \ldots, T_{l-1}, T_{l+1}, \ldots, T_k \text{ are } \leq t)-P_{\underset{\sim}{\theta}}(\text{At least } i \text{ of } T_1, \ldots,$$

$$T_{l-1}, T_{l+1}, \ldots, T_k \text{ are } \leq t)\}+P_{\underset{\sim}{\theta}}(\text{At least } i \text{ of } T_1, \ldots, T_{l-1}, T_{l+1}, \ldots, T_k \text{ are } \leq t)$$

which is a non-increasing function of θ_l because (i) $\{\ \}$ is nonnegative, (ii) $P_{\underset{\sim}{\theta}}$ (At least i of $T_1, \ldots, T_{l-1}, T_{l+1}, \ldots, T_k$ are $\leq t$) does not involve θ_l, and (iii) $G_n(t; \theta_l)$ is a non-increasing function of θ_l by assumption.

ASSUMPTION (2.2): $G_n(t; \theta)$ is said to be degenerate at θ_* and θ^* if

(2.2a) (2.2b)
$$G_n(t; \theta)=P_\theta(T\leq t)=\begin{cases}1 & \text{for } \theta=\theta_* \\ 0 & \text{for } \theta=\theta^*\end{cases}$$

where θ_* and θ^* are the smallest and the largest possible values of θ respectively and $t\epsilon(T_*, T^*)$.

In the following, we will consider a random interval I for $\theta_{[i]}(1\leq i\leq k)$. For preassigned $\gamma\epsilon(0, 1)$, we say the event CD ("Correct Decision") occurs iff for any $i(1\leq i\leq k)\theta_{[i]}\epsilon I$. One usually tries to develop a procedure R for the construction of different I's in such a way that

(2.3)
$$\inf_{\underset{\sim}{\theta}\epsilon\Omega(\theta_{[i]})} P_{\underset{\sim}{\theta}}(CD|R)\geq\gamma$$

where

$$\Omega(\underset{\sim}{\theta}_{[i]})=\{\underset{\sim}{\theta}=(\theta_{[i]}, \ldots, \theta_{[k]})|\theta_{[i]} \text{ is held fixed}\}.$$

A. Lower Confidence Intervals

PROCEDURE R_L: Let h_1 be a continuous increasing function with inverse g_2. Define

(2.4)
$$I_L=(h_1(T_{[i]}|\gamma_1), \theta^*)$$

to be a lower confidence interval for $\theta_{[i]}$ with coverage probability γ_1. We have CD iff $\theta_{[i]}\epsilon I_L$.

THEOREM (2.5): Under procedure R_L, if Assumption (2.2b) holds, then

$$\inf_{\underset{\sim}{\theta}\epsilon\Omega(\theta_{[i]})} P_{\underset{\sim}{\theta}}(CD|R_L)=\{G_n(g_2(\theta_{[i]}); \theta_{[i]})\}^i,$$

and given $\gamma_1\ \epsilon\ (0, 1)$, k, n, and i, $(1\leq i\leq k)$, the end point $h_1(T_{[i]}|\gamma_1)$ can be obtained as the unique solution in $\theta_{[i]}$ of

(2.6)
$$G_n(g_2(\theta_{[i]}); \theta_{[i]})=G_n(T_{[i]}; \theta_{[i]})=\gamma_1^{1/i}.$$

PROOF: By Lemma (2.1) and Assumption (2.2b), we have

$$P_{\underset{\sim}{\theta}}(CD|R)=P_{\underset{\sim}{\theta}}(T_{[i]}\leq g_2(\theta_{[i]}))\geq P_{\underset{\sim}{\theta}}(T_{[i]}\leq g_2(\theta_{[i]})|\underset{\sim}{\theta}=(\underbrace{\theta_{[1]}, \ldots, \theta_{[i]}}_{i}, \underbrace{\theta^*, \ldots, \theta^*}_{k-i}))$$

$$=P_{\theta_{[i]}}(\max(Y_1, \ldots, Y_i)\leq g_2(\theta_{[i]}))$$

$$=\{G_n(g_2(\theta_{[i]}); \theta_{[i]})\}^i$$

where $Y_1, \ldots, Y_i$ are i.i.d. r.v.'s with c.d.f. $G_n(t; \theta_{[i]})$. The rest of the proof of this theorem is clear.

PROCEDURE R_U: Let h_2 be a continuous increasing function with inverse g_1. Define

$$I_U=(\theta_*, h_2(T_{[i]}|\gamma_2)) \tag{2.7}$$

to be an upper confidence interval for $\theta_{[i]}$ with coverage probability γ_2. We have CD iff $\theta_{[i]}\epsilon I_U$.

THEOREM (2.8): Under procedure R_U, if Assumption (2.2a) holds, then

$$\operatorname*{Inf}_{\underset{\sim}{\theta}\epsilon\Omega(\theta_{[i]})} P_{\underset{\sim}{\theta}}(CD|R_U)=\{1-G_n(g_1(\theta_{[i]}); \theta_{[i]})\}^{k-i+1}$$

and given $\gamma_2\epsilon$ (0, 1), k, n, and i $(1\leq i\leq k)$, the end point $h_2(T_{[i]}|\gamma_2)$ can be obtained as the unique solutions in $\theta_{[i]}$ of

$$G_n(g_1(\theta_{[i]}); \theta_{[i]})=G_n(T_{[i]}; \theta_{[i]})=1-\gamma_2^{1/(k-i+1)}. \tag{2.9}$$

PROOF: By Lemma (2.1) and Assumption (2.2a), the proof is similar to that of Theorem (2.5) and will be omitted.

B. Two-Sided Confidence Intervals

In Part A we have constructed a lower and an upper confidence interval for $\theta_{[i]}$ with coverage probability at least γ_1 and γ_2 $(0<\gamma_1, \gamma_2<1)$ each respectively. Here we will adopt the method used by Dudewicz [5] and obtain a class of two-sided confidence intervals on $\theta_{[i]}$ $(1\leq i\leq k)$ if assumption (2.2) holds.

PROCEDURE R_T: Let

$$I_T=(h_1(T_{[i]}|\gamma_1), h_2(T_{[i]}|\gamma_2)) \tag{2.10}$$

be the intersection of the lower interval I_L and the upper interval I_U with coverage probabilities at least γ_1 and γ_2 respectively. We have CD iff $\theta_{[i]}\epsilon I_T$.

THEOREM (2.11): Let γ_1, γ_2 $(0<\gamma_1, \gamma_2<1)$ be such that $\gamma_1+\gamma_2-1=\alpha$ (α fixed and $0<\alpha<1$). Then, under R_T, the interval I_T is a two-sided confidence interval for $\theta_{[i]}$ with coverage probability at least α.

PROOF: Define $A=\{\theta_{[i]} \geq h_1(T_{[i]}|\gamma_1)\}$ and $B=\{\theta_{[i]} \leq h_2(T_{[i]}|\gamma_2)\}$, one intercepts the lower and the upper intervals and proceeds as in the same proof used by Dudewicz [5].

The length of I_T is given by

$$L(\gamma_1)=h_2(T_{[i]}|\alpha+1-\gamma_1)-h_1(T_{[i]}|\gamma_1). \tag{2.12}$$

The optimum choice of γ_1 which minimizes the length $L(\gamma_1)$ can be determined by numerical methods.

Alam and Saxena [1] have also constructed a two-sided confidence interval for $\theta_{[i]}$†. They started from the general formulation of two-sided interval by using the incomplete beta function. While our approach is to intercept two one-sided intervals to obtain the desired two-sided intervals. When Assumption (2.2), holds our results and theirs coincide. Assumption (2.2) is basic to our work. For example, when T is a normal r.v. with mean μ and variance one, then $P(T\leq t)=\Phi(t-\mu)$ approaches 1 as $\mu\to-\infty$ and 0 as $\mu\to+\infty$, where $\Phi(.)$ is the c.d.f. of a standard normal r.v. As a second example, when T is a Chi-square r.v. time σ^2, then $P(T\leq t)=G(t/\sigma^2)$ approaches 1 as $\sigma^2\to 0$ and 0 as $\sigma^2\to\infty$, where $G(.)$ is the c.d.f. of a Chi-square r.v. Theorems (2.5) and (2.8) are based on this property.

THEOREM (2.13): For any i $(1\leq i\leq k)$, the lower confidence interval I_L on $\theta_{[i]}$ which has minimal coverage probability γ_1 has maximal coverage probability $1-(1-\gamma_1^{1/i})^{k-i+1}$. And the upper confidence interval I_U on $\theta_{[i]}$ which has minimal coverage probability γ_2 has maximal coverage probability $1-(1-\gamma_2^{1/(k-i+1)})^i$.

PROOF OF THE LOWER CONFIDENCE INTERVAL CASE: Based on Assumption (2.2) we have

$$\begin{aligned} P_{\underset{\sim}{\theta}}(T_{[i]}\leq g_2(\theta_{[i]})) &\leq P(T_{[i]}\leq g_2(\theta_{[i]})|\underset{\sim}{\theta}=(\underbrace{\theta_*,\ldots,\theta_*}_{i-1},\underbrace{\theta_{[i]},\ldots,\theta_{[i]}}_{k-i+1}) \\ &=P_{\theta_{[i]}}(\min(Y_1,\ldots,Y_{k-i+1})\leq g_2(\theta_{[i]})) \\ &=1-\{1-P_{\theta_{[i]}}(\min(Y_1,\ldots,Y_{k-i+1})\leq g_2(\theta_{[i]}))\} \end{aligned}$$

where

$Y_1,\ldots,Y_{k-i+1}$ are i.i.d. r.v.'s with c.d.f. $G_n(t;\theta_{[i]})$. Hence

$$\begin{aligned} \sup_{\underset{\sim}{\theta}\in\Omega(\theta_{[i]})} P_{\underset{\sim}{\theta}}(T_{[i]}\leq g_2(\theta_{[i]})) &= 1-\{1-P_{\theta_{[i]}}(\min(Y_1,\ldots,Y_{k-i+1})\leq g_2(\theta_{[i]}))\} \\ &= 1-\{1-G_n(g_2(\theta_{[i]});\theta_{[i]})\}^{k-i+1}=1-\{1-\gamma_1^{1/i}\}^{k-i+1} \end{aligned}$$

where

$$G_n(g_2(\theta_{[i]});\theta_{[i]})=\gamma_1^{1/i} \text{ by (2.6)}.$$

The proof of the upper confidence interval case is similar to that of the lower one and will be omitted.

---

†After these results were obtained, the author received a manuscript of Alam and Saxena [1] which independently developed similar results on interval estimation of a ranked parameter.

Table (2.14) of $1-(1-\gamma_1^{1/i})^{k-i+1}$ illustrates the maximal degree of overprotection when $k=4$ and 5.

Table (2.14). $1-(1-\gamma_1^{1/i})^{k-i+1}$

k	$k=4$				$k=5$				
γ \ i	$i=1$	$i=2$	$i=3$	$i=4$	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$
0. 99	1. 000	1. 000	1. 000	0. 997	1. 000	1. 000	1. 000	1. 000	0. 998
0. 95	1. 000	1. 000	1. 000	0. 987	1. 000	1. 000	1. 000	1. 000	0. 990
0. 90	1. 000	1. 000	0. 999	0. 974	1. 000	1. 000	1. 000	0. 999	0. 979
0. 80	0. 998	0. 999	0. 995	0. 946	1. 000	1. 000	1. 000	0. 997	0. 956
0. 70	0. 992	0. 996	0. 987	0. 915	0. 998	0. 999	0. 999	0. 993	0. 931
0. 60	0. 974	0. 989	0. 975	0. 880	0. 990	0. 997	0. 996	0. 986	0. 903
0. 50	0. 938	0. 975	0. 957	0. 841	0. 969	0. 993	0. 991	0. 975	0. 871
0. 30	0. 760	0. 907	0. 891	0. 740	0. 832	0. 958	0. 964	0. 932	0. 786
0. 10	0. 344	0. 680	0. 713	0. 562	0. 410	0. 781	0. 846	0. 808	0. 631

LEMMA (2.15): If X and Y are independent r.v.'s with $F_X(x)=P(X\leq x)\leq P(Y\leq x)=F_Y(x)$, $x\epsilon R$, then $E(X)\geq E(Y)$.

THEOREM (2.16): For any $i(1\leq i\leq k)$, by Lemma (2.1)

$$\text{Inf}\ \{E_{\underset{\sim}{\theta}}(T_{[i]});\ \theta\epsilon\Omega(\theta_{[i]})\}=E(i^{\text{th}}\text{ smallest of } T_1,\ \ldots,\ T_k),$$

$$\underset{\sim}{\theta}=(\underbrace{\theta_*,\ \ldots,\ \theta_*}_{i-1},\ \underbrace{\theta_{[i]},\ \ldots,\ \theta_{[i]}}_{k-i+1})$$

and

$$\text{Sup}\ \{E_{\underset{\sim}{\theta}}(T_{[i]});\ \theta\epsilon\Omega(\theta_{[i]})\}=E(i^{\text{th}}\text{ smallest of } T_1,\ \ldots,\ T_k).$$

$$\underset{\sim}{\theta}=(\underbrace{\theta_{[i]},\ \ldots,\ \theta_{[i]}}_{i},\ \underbrace{\theta^*,\ \ldots,\ \theta^*}_{k-i})$$

PROOF OF THE INFIMUM CASE: Fix any $i(1\leq i\leq k)$; by Lemmas (2.1) and (2.15) we have

$$E_{\underset{\sim}{\theta}}(T_{[i]})=E_{\underset{\sim}{\theta}}(i^{\text{th}}\text{ smallest of } T_1,\ \ldots,\ T_k)\geq E(i^{\text{th}}\text{ smallest of } T_1,\ \ldots,\ T_k)$$

$$\underset{\sim}{\theta}=(\underbrace{\theta_*,\ \ldots,\ \theta_*}_{i-1},\ \underbrace{\theta_{[i]},\ \ldots,\ \theta_{[i]}}_{k-i+1})$$

which completes the proof. The proof of supremum case is similar to that of infimum one and will be omitted.

If Assumption (2.2) holds then

$$\text{Inf}\ \{E_{\underset{\sim}{\theta}}(T_{[i]});\ \underset{\sim}{\theta}\epsilon\Omega(\theta_{[i]})\}=E_{\theta_{[i]}}(\min\ (T_1,\ \ldots,\ T_{k-i+1})),$$

and

$$\text{Sup}\ \{E_{\underset{\sim}{\theta}}(T_{[i]});\ \underset{\sim}{\theta}\epsilon\Omega(\theta_{[i]})\}=E_{\theta_{[i]}}(\max\ (T_1,\ \ldots,\ T_i))$$

where $T_1,\ \ldots,\ T_{k-i+1}$ or $T_1,\ \ldots,\ T_i$ are i.i.d. r.v.'s with parameter $\theta_{[i]}$.

3. SOME APPLICATIONS OF A RANKED PARAMETER

A. Location Parameter

Suppose that $G_n(t;\theta_i)$ is a continuous c.d.f. and $G_n(t;\theta_i)=G_n(t-\theta_i)$ for $t\epsilon(T_*, T^*)$ and $\theta_i\epsilon(\theta_*, \theta^*)$, $i=1, \ldots, k$. Then, from (2.4), (2.6), (2.7), (2.9), (2.10), and Theorem (2.11), the lower, the upper, and the two-sided confidence intervals for $\theta_{[i]}$ with coverage probabilities γ_1, γ_2, and $\gamma_1+\gamma_2-1=\alpha$ respectively are given by

$$I_L=(T_{[i]}-G_n^{-1}(\gamma_1^{1/i}), \theta^*),$$

$$I_U=(\theta_*, T_{[i]}-G_n^{-1}(1-\gamma_2^{1/(k-i+1)})),$$

and

$$I_T=(T_{[i]}-G_n^{-1}(\gamma_1^{1/i}), T_{[i]}-G_n(1-\gamma_2^{1/(k-i+1)}))$$

where $G_n^{-1}(.)$ is the inverse of $G_n(.)$.

For normal populations with know variances the confidence intervals, with $T_{[i]}=\overline{X}_{[i]}$ and $G_n(.)=\Phi(.)$, for $\theta_{[i]}=\mu_{[i]}$ are the results of Dudewicz [4, 5, 6].

B. Scale Parameter

Suppose that $G_n(t;\theta_i)$ is a continuous c.d.f. and $G_n(t;\theta_i)=G_n(t/\theta_i)$ for all $t\epsilon(T_*, T^*)$ and $\theta_i\epsilon(\theta_*, \theta^*)$, $i=1, \ldots, k$. Then, from (2.4), (2.6), (2.7), (2.9), (2.10), and Theorem (2.11), the lower, the upper, and the two-sided confidence intervals for $\theta_{[i]}$ with coverage probabilities γ_1, γ_2, and $\gamma_1+\gamma_2-1=\alpha$ respectively are given by

$$I_L=(T_{[i]}/G_n^{-1}(\gamma_1^{1/i}), \theta^*)$$

$$I_U=(\theta_*, T_{[i]}/G_n^{-1}(1-\gamma_2^{1/(k-i+1)})),$$

and

$$I_T=(T_{[i]}/G_n^{-1}(\gamma_1^{1/i}), T_{[i]}/G_n^{-1}(1-\gamma_2^{1/(k-i+1)})).$$

In the following we give an example of a normal variances case.

Let observations from $\pi_i(1\leq i\leq k)$ be independent and normally distributed $N(\mu_i, \sigma_i^2)$ with scale parameter σ_i^2 unknown. The goal is to estimate $\sigma_{[i]}^2$, the i^{th} smallest variance. Let $S_1^2, \ldots, S_k^2$ be mutually independent unbiased variance estimators, each based on size n for $\sigma_1^2, \ldots, \sigma_k^2$ from populations $\pi_1, \ldots, \pi_k$ respectively. We know that S_i^2 has σ_i^2 as a scale parameter with p.d.f. $g_n(x;\sigma_i^2)$ and c.d.f. $G_n(x;\sigma_i^2)$ which depend on x, σ_i^2 only through x/σ_i^2, $i=1, \ldots, k$. $g_n(x)=g_n(x;1)$ and $G_n(x)=G_n(x;1)$ are respectively the p.d.f. and c.d.f. of a chi-square/$(n-1)$ r.v. with $(n-1)$ d.f.

The theorems and lemmas of Section 2 can be applied to the above normal variances case as follows:

(1). For any i $(1\leq i\leq k)$, the c.d.f. of $S_{[i]}^2$, namely $H_{S_{[i]}^2}(x)$, decreases as σ_l^2 $(1\leq l\leq k)$ increases;

(2). For any i $(1\leq i\leq k)$,

$$\text{Inf}\{E_{\underset{\sim}{\sigma}^2}(S_{[i]}^2); \underset{\sim}{\sigma}^2\in\Omega(\sigma_{[i]}^2)\}=\sigma_{[i]}^2 h'_{k-i+1}(g_n) \tag{3.1}$$

and

$$\text{Sup}\ \{E_{\underset{\sim}{\sigma}^2}(S^2_{[i]});\ \underset{\sim}{\sigma}^2 \in \Omega(\sigma^2_{[i]})\} = \sigma^2_{[i]} h_i(g_n) \tag{3.2}$$

where

$$h'_{k-i+1}(g_n) = \int_0^\infty y(k-i+1)[1-G_n(y)]^{k-i} g_n(y)\,dy$$

and

$$h'_i(g_n) = \int_0^\infty yi\,[G_n(y)]^{i-1} g_n(y)\,dy;$$

(3). For any i $(1 \le i \le k)$,

$$\sigma^2_{[i]} h'_{k-1+i}(g_n) \le E_{\underset{\sim}{\sigma}^2}(S^2_{[i]}) \le \sigma^2_{[i]} h_i(g_n); \tag{3.3}$$

(4). For any i $(1 \le i \le k)$, an upper confidence interval for $\sigma^2_{[i]}$ with coverage probability at least γ_2 is given by $(0,\ S^2_{[i]}/b_n)$, where b_n is the solution of

$$(1-G_n(b_n))^{k-i+1} = \gamma_2 \tag{3.4}$$

where $G_n(y)$ is the c.d.f. of a chi-square/$(n-1)$ r.v. with $n-1$ d.f.;

(5). For any i $(1 \le i \le k)$, a lower confidence interval for $\sigma^2_{[i]}$ with coverage probability at least γ_1 is given by $(S^2_{[i]}/a_n,\ \infty)$ where a_n is the solution of

$$G_n(a_n)^i = \gamma_1; \tag{3.5}$$

and

(6). For any i $(1 \le i \le k)$, the upper confidence interval of (4) on $\sigma^2_{[i]}$ which has minimal probability of coverage γ_2 has maximal probability of coverage $1-(1-\gamma_2^{1/(k-i+1)})^i$. And the lower confidence interval of (5) on $\sigma^2_{[i]}$ which has minimal probability of coverage γ_1 has maximal probability of coverage $1-(1-\gamma_1^{1/i})^{k-i+1}$;

(7). In (4) and (5) we have obtained an upper and a lower confidence interval for $\sigma^2_{[i]}$ with coverage probability at least γ_2 and γ_1 $(0<\gamma_1,\ \gamma_2<1)$ each respectively. We now obtain a class of two-sided confidence intervals for $\sigma^2_{[i]}$ $(1 \le i \le k)$ by Theorem (2.11), the intervals are of the form

$$(S^2_{[i]}/G_n^{-1}(\gamma_1^{1/i}),\ S^2_{[i]}/G_n^{-1}(1-\gamma_2^{1/(k-i+1)}))$$

being a class of two-sided confidence intervals for $\sigma^2_{[i]}$, where $G_n^{-1}(\cdot)$ is the inverse of the c.d.f. $G_n(\cdot)$ of a chi-square/$(n-1)$ r.v. with $(n-1)$ d.f.

Tables for values of b_n and a_n are given in Tables (3.6) and (3.7) for any i $(1 \le i \le k)$ and any k $(1 \le k \le 5)$. For example, given $\gamma=0.95$, $k=5$, $i=3$, and d.f.$=30$, we wish to find the coefficient b_n of an upper confidence interval on $\sigma^2_{[3]}$, we have $k-i+1=5-3+1=3$, read Table (3.6) and find the column with number 3 on the top then we have $b_n=0.5321$ with d.f.$=30$; the coefficient a_n of a lower confidence interval on $\sigma^2_{[3]}$ can be found by reading the column 3 of Table (3.7), we obtain $a_n=1.6226$ with d.f.$=30$. The figures of a_n and b_n are accurate up to the fourth decimal place (with ± 1 unit) according to the error analysis of Dudewicz, Ramberg, and Chen [7].

TABLE (3.6). Values of b_n for $1 \leq i \leq k (1 \leq k \leq 5)$.

	$\gamma=0.95$					$\gamma=0.99$				
df \ $k-i+1$	5	4	3	2	1	5	4	3	2	1
1	0. $0^{3}16$	0. $0^{3}26$	0. $0^{3}45$	0. 0010	0. 0039	0. $0^{5}60$	0. $0^{4}10$	0. $0^{4}18$	0. $0^{4}39$	0. $0^{3}16$
2	0. 0103	0. 0128	0. 0171	0. 0256	0. 0513	0. 0020	0. 0025	0. 0033	0. 0050	0. 0101
3	0. 0388	0. 0452	0. 0550	0. 0726	0. 1173	0. 0129	0. 0150	0. 0182	0. 0239	0. 0383
4	0. 0751	0. 0844	0. 0982	0. 1219	0. 1777	0. 0324	0. 0363	0. 0421	0. 0518	0. 0743
5	0. 1118	0. 1232	0. 1397	0. 1672	0. 2291	0. 0561	0. 0616	0. 0695	0. 0824	0. 1109
6	0. 1465	0. 1592	0. 1774	0. 2073	0. 2726	0. 0812	0. 0879	0. 0974	0. 1127	0. 1454
7	0. 1782	0. 1918	0. 2112	0. 2425	0. 3096	0. 1060	0. 1136	0. 1244	0. 1414	0. 1770
8	0. 2071	0. 2213	0. 2414	0. 2736	0. 3416	0. 1298	0. 1382	0. 1498	0. 1682	0. 2058
9	0. 2333	0. 2479	0. 2685	0. 3012	0. 3695	0. 1524	0. 1613	0. 1735	0. 1929	0. 2320
10	0. 2571	0. 2720	0. 2929	0. 3258	0. 3940	0. 1736	0. 1829	0. 1958	0. 2157	0. 2558
11	0. 2789	0. 2940	0. 3150	0. 3480	0. 4159	0. 1935	0. 2031	0. 2163	0. 2368	0. 2776
12	0. 2989	0. 3141	0. 3351	0. 3681	0. 4355	0. 2121	0. 2219	0. 2355	0. 2563	0. 2976
13	0. 3173	0. 3325	0. 3535	0. 3864	0. 4532	0. 2295	0. 2395	0. 2533	0. 2744	0. 3159
14	0. 3342	0. 3494	0. 3704	0. 4032	0. 4693	0. 2459	0. 2560	0. 2699	0. 2912	0. 3329
15	0. 3500	0. 3651	0. 3861	0. 4186	0. 4841	0. 2612	0. 2715	0. 2855	0. 3069	0. 3486
16	0. 3646	0. 3797	0. 4005	0. 4328	0. 4976	0. 2757	0. 2860	0. 3001	0. 3215	0. 3633
17	0. 3783	0. 3933	0. 4140	0. 4460	0. 5101	0. 2893	0. 2997	0. 3138	0. 3353	0. 3769
18	0. 3911	0. 4060	0. 4266	0. 4583	0. 5217	0. 3022	0. 3126	0. 3267	0. 3482	0. 3897
19	0. 4030	0. 4179	0. 4383	0. 4698	0. 5325	0. 3144	0. 3248	0. 3389	0. 3603	0. 4017
20	0. 4143	0. 4291	0. 4494	0. 4806	0. 5425	0. 3259	0. 3363	0. 3504	0. 3718	0. 4130
25	0. 4622	0. 4765	0. 4960	0. 5258	0. 5845	0. 3757	0. 3860	0. 4000	0. 4209	0. 4610
30	0. 4997	0. 5134	0. 5321	0. 5606	0. 6164	0. 4156	0. 4257	0. 4393	0. 4597	0. 4984
40	0. 5553	0. 5680	0. 5854	0. 6117	0. 6627	0. 4760	0. 4856	0. 4985	0. 5178	0. 5541
60	0. 6258	0. 6371	0. 6524	0. 6755	0. 7198	0. 5546	0. 5633	0. 5750	0. 5924	0. 6247
100	0. 7015	0. 7110	0. 7237	0. 7429	0. 7793	0. 6412	0. 6487	0. 6587	0. 6734	0. 7007

Note that $0.0^{3}16$ means 0.00016 and $0.0^{5}60$ means 0.0000060.

C. Binomial Parameter

A binomial distribution belongs to the family of monotone likelihood ratio which is also stochastically increasing in parameter (see, e.g., Lehmann [9]) and satisfies Assumption (2.2). Take n independent observations $X_{i1}, \ldots, X_{in}$ from Bernoulli population π_i with proportion (parameter) $\theta_i=P_i$ $(0 \leq P_i \leq 1)$, $i=1, \ldots, k$, then the sample mean

$$\overline{X}_i=\sum_{j=1}^{n} X_{ij}/n,$$

as an estimator of P_i, is a binomial r.v. with parameter P_i, $i=1, \ldots, k$. Let $\overline{X}_{[1]} \leq \ldots \leq \overline{X}_{[k]}$ denote the ranked values of $\overline{X}_1, \ldots, \overline{X}_k$, and $P_{[1]} \leq \ldots \leq P_{[k]}$ the ranked values of $P_1, \ldots, P_k$. The c.d.f. of $\overline{X}_i$ is given by

$$G_n(x; P_i)=1-I_{P_i}([nx]+1, n-[nx])$$

where $I_{P_i}(.)$ is the incomplete beta function and $[y]$ is the smallest integer $\geq y$. We note that the observed value of $n\overline{X}_{[i]}$ will be an integer. Then, from (2.6) and (2.9), the end points $h_1(\overline{X}_{[i]}|\gamma_1)$

TABLE (3.7). Values of a_n for $1 \leq i \leq k$ $(1 \leq k \leq 5)$.

	$\gamma=0.95$					$\gamma=0.99$				
df \ i	1	2	3	4	5	1	2	3	4	5
1	3. 8415	5. 0018	5. 7013	6. 2047	6. 5985	6. 6349	7. 8748	8. 6093	9. 1335	9. 5419
2	2. 9957	3. 6761	4. 0773	4. 3629	4. 5848	4. 6052	5. 2958	5. 7005	5. 9878	6. 2105
3	2. 6049	3. 1068	3. 3995	3. 6066	3. 7669	3. 7816	4. 2776	4. 5664	4. 7708	4. 9290
4	2. 3719	2. 7783	3. 0136	3. 1794	3. 3074	3. 3192	3. 7137	3. 9424	4. 1039	4. 2287
5	2. 2141	2. 5601	2. 7594	2. 8994	3. 0074	3. 0173	3. 3487	3. 5403	3. 6754	3. 7796
6	2. 0986	2. 4026	2. 5770	2. 6993	2. 7934	2. 8020	3. 0902	3. 2564	3. 3734	3. 4636
7	2. 0096	2. 2825	2. 4385	2. 5477	2. 6317	2. 6393	2. 8959	3. 0435	3. 1473	3. 2273
8	1. 9384	2. 1873	2. 3291	2. 4282	2. 5043	2. 5113	2. 7435	2. 8769	2. 9706	3. 0428
9	1. 8799	2. 1095	2. 2400	2. 3311	2. 4009	2. 4073	2. 6203	2. 7424	2. 8280	2. 8940
10	1. 8307	2. 0444	2. 1657	2. 2502	2. 3150	2. 3209	2. 5181	2. 6310	2. 7102	2. 7711
11	1. 7886	1. 9891	2. 1026	2. 1816	2. 2422	2. 2477	2. 4318	2. 5371	2. 6108	2. 6675
12	1. 7522	1. 9413	2. 0482	2. 1226	2. 1795	2. 1848	2. 3577	2. 4565	2. 5257	2. 5788
13	1. 7502	1. 8995	2. 0007	2. 0711	2. 1249	2. 1299	2. 2932	2. 3865	2. 4517	2. 5018
14	1. 6918	1. 8625	1. 9588	2. 0257	2. 0769	2. 0815	2. 2366	2. 3249	2. 3868	2. 4342
15	1. 6664	1. 8296	1. 9215	1. 9853	2. 0341	2. 0385	2. 1862	2. 2704	2. 3292	2. 3744
16	1. 6435	1. 8000	1. 8880	1. 9491	1. 9957	2. 0000	2. 1412	2. 2216	2. 2778	2. 3209
17	1. 6228	1. 7732	1. 8577	1. 9164	1. 9611	1. 9652	2. 1006	2. 1776	2. 2314	2. 2727
18	1. 6038	1. 7488	1. 8302	1. 8866	1. 9297	1. 9336	2. 0638	2. 1378	2. 1894	2. 2291
19	1. 5865	1. 7265	1. 8051	1. 8595	1. 9010	1. 9048	2. 0302	2. 1015	2. 1512	2. 1893
20	1. 5705	1. 7060	1. 7820	1. 8345	1. 8747	1. 8783	1. 9994	1. 0682	2. 1162	2. 1529
25	1. 5061	1. 6237	1. 6894	1. 7348	1. 7694	1. 7726	1. 8767	1. 9358	1. 9769	2. 0084
30	1. 4591	1. 5641	1. 6226	1. 6629	1. 6936	1. 6964	1. 7887	1. 8410	1. 8773	1. 9051
40	1. 3940	1. 4820	1. 5308	1. 5644	1. 5899	1. 5923	1. 6689	1. 7121	1. 7421	1. 7650
60	1. 3180	1. 3871	1. 4252	1. 4513	1. 4712	1. 4730	1. 5323	1. 5657	1. 5888	1. 6064
100	1. 2334	1. 2947	1. 3229	1. 3421	1. 3567	1. 3581	1. 4015	1. 4259	1. 4427	1. 4555

and $h_2(\overline{X}_{[i]}|\gamma_2)$ of the intervals given by (2.4), (2.7), and (2.10) are determined by the unique solutions in $P_{[i]}$ of the following equations respectively

$$1-I_{P_{[i]}}(n\overline{X}_{[i]}+1,\ n-n\overline{X}_{[i]})=\gamma_1^{1/i}, \tag{3.8}$$

$$1-I_{P_{[i]}}(n\overline{X}_{[i]},\ n-n\overline{X}_{[i]}+1)=1-\gamma_2^{1/(k-i+1)}. \tag{3.9}$$

The end points in (3.8) and (3.9) can be obtained by numerical methods from the incomplete beta function tables (see, e.g., K. Pearson [12]) and its interpolation.

LEMMA (3.10): If $X_1, \ldots, X_i$ are independent r.v.'s each having Binomial distribution $(n, P_{[i]})$, and

$$\overline{X}_{[i]}=\max\{\overline{X}_1, \ldots, \overline{X}_i\},\ \overline{X}_{[1]}=\min\{\overline{X}_1, \ldots, \overline{X}_i\}$$

where $\overline{X}_i=X_i/n$, and $X_i=\sum_{j=1}^{n}X_{ij}$, then

$$E(\overline{X}_{[i]})=1-\frac{1}{n}\sum_{x=0}^{n-1}(\Pr(x))^i$$

$$E(\overline{X}_{[1]})=\frac{1}{n}\sum_{x=0}^{n-1}(1-\Pr(x))^i$$

where

$$\Pr(t)=\sum_{\alpha=0}^{t}\binom{n}{\alpha}P^{\alpha}_{[i]}(1-P_{[i]})^{n-\alpha}$$

$$=1-I_{P_{[i]}}(t+1, n-t) \qquad t=0, 1, 2, \ldots, n.$$

PROOF: By the property of $P(X_{[1]}\leq x)=1-(1-Pr(x))^{l}$ and $P(X_{[l]}\leq x)=(Pr(x))^{l}$, and using the result on p. 211 of Parzen (1960) we have

$$E(X_{[1]})=\int_0^{\infty}(1-\Pr(x))^{l}dx=\sum_{x=0}^{n-1}(1-\Pr(x))^{l}$$

and

$$E(X_{[l]})=\int_0^{\infty}(1-(\Pr(x))^{l})dx=\sum_{x=0}^{n-1}(1-(\Pr(x))^{l})$$

LEMMA (3.11): For any i $(1\leq i\leq k)$,

$$\operatorname*{Sup}_{\underset{\sim}{P}\in\Omega(P_{[i]})}E_{\underset{\sim}{P}}(\overline{X}_{[i]})=\frac{1}{n}\sum_{u=0}^{n-1}\{1-(\Pr(u))^{i}\}$$

and

$$\operatorname*{Inf}_{\underset{\sim}{P}\in\Omega(P_{[i]})}E_{\underset{\sim}{P}}(\overline{X}_{[i]})=\frac{1}{n}\sum_{u=0}^{n-1}\{1-(\Pr(u))^{k-i+1}\}$$

PROOF: This follows from Theorem (2.16) and Lemma (3.10).

LEMMA (3.12): For any $i(1\leq i\leq k)$

$$\lim_{P_{[i]}\to 0}\operatorname*{Sup}_{\underset{\sim}{P}}E_{\underset{\sim}{P}}(\overline{X}_{[i]})=0=\lim_{P_{[i]}\to 0}\operatorname*{Inf}_{\underset{\sim}{P}}E_{\underset{\sim}{P}}(\overline{X}_{[i]})$$

$$\lim_{P_{[i]}\to 1}\operatorname*{Sup}_{\underset{\sim}{P}}E_{\underset{\sim}{P}}(\overline{X}_{[i]})=1=\lim_{P_{[i]}\to 1}\operatorname*{Inf}_{\underset{\sim}{P}}E_{\underset{\sim}{P}}(\overline{X}_{[i]}).$$

PROOF: Since $\Pr(u)$ approaches 1(0) as P approaches 0(1) and by using Lemma (3.11). From Theorem (2.16) and Lemma (3.11) we have

$$\frac{1}{n}\sum_{u=0}^{n-1}\{(1-\Pr(u))^{k-i+1}\}\leq E_{\underset{\sim}{P}}(\overline{X}_{[i]})\leq\frac{1}{n}\sum_{u=0}^{n-1}\{1-(\Pr(u))^{i}\}.$$

D. Poisson Parameter

The Poisson distribution with parameter λ belongs to the family of monotone likelihood ratio which is stochastically increasing in λ. Take n independent observations $X_{i1}, \ldots, X_{in}$ from Poisson population π_i with parameter $\lambda_i(0<\lambda_i<\infty)$, $i=1, \ldots, k$, then the sample mean $\overline{X}_i=\sum_{j=1}^{k}X_{ij}/n$, as an estimator of λ_i, is again a Poisson r.v. with parameter $n\lambda_i$, $i=1, \ldots, k$. Let $\overline{X}_{[i]}$ denote the i^{th} smallest value of $\overline{X}$'s and $\lambda_{[i]}$ the i^{th} smallest value of λ's, $i=1, \ldots, k$. The c.d.f. of $\overline{X}_i$ is given by

$$G_n(x;\lambda_i)=1-\Gamma_{n\lambda i}([nx]+1)$$

where $\Gamma_{n\lambda_i}(.)$ is the incomplete gamma function. Then, from (2.6) and (2.9), the end points of the intervals given by (2.4), (2.7), and (2.10) are determined by the unique solutions in $n\lambda_{[i]}$ of the following equations respectively

$$1-\Gamma_{n\lambda_{[i]}}(n\overline{X}_{[i]}+1)=\gamma_1^{1/t}, \tag{3.13}$$

$$1-\Gamma_{n\lambda_{[i]}}(n\overline{X}_{[i]})=1-\gamma_2^{1/(k-i+1)}. \tag{3.14}$$

The end points in (3.13) and (3.14) can be obtained by numerical methods from the incomplete gamma function tables (see, e.g., K. Pearson [11]) and its interpolation. From Theorem (2.16) and a lemma similar to Lemma (3.10) we have

$$\text{Sup}\{E_{\underset{\sim}{\lambda}}(\overline{X}_{[i]});\ \underset{\sim}{\lambda}\in\Omega(\lambda_{[i]})\}=E_{\lambda_{[i]}}\{\text{max of } \overline{X}_{(1)},\ldots,\overline{X}_{(i)}\}$$

$$=\frac{1}{n}\sum_{u=0}^{n-1}\{1-(\Pr(u))^i\}$$

$$\text{Inf}\{E_{\underset{\sim}{\lambda}}(\overline{X}_{[i]});\ \underset{\sim}{\lambda}\in\Omega(\lambda_{[i]})\}=E_{\lambda_{[i]}}\{\text{min of } \overline{X}_{(i)},\ldots,\overline{X}_{(k)}\}$$

$$=\frac{1}{n}\sum_{u=0}^{n-1}\{(1-\Pr(u))^{k-i+1}\}$$

where

$$\Pr(u)=1-\Gamma_{n\lambda_{[i]}}(u+1)=\sum_{\alpha=0}^{u}(e^{-n\lambda_{[i]}}(n\lambda_{[i]})^\alpha)/\alpha!,\ u=0,1,2,\ldots$$

E. Other Examples

Other examples include the noncentrality parameter of a noncentral t distribution and the correlation coefficient of a bivariate normal distribution.

ACKNOWLEDGMENTS

The author wishes to thank Professor Edward J. Dudewicz and a referee for comments and suggestions on earlier version of this paper.

BIBLIOGRAPHY

[1] Alam, K. and K. M. L. Saxena, "On Interval Estimation of a Ranked Parameter," Abstract, Bulletin, Institute of Mathematical Statistics, *2* (3) 118 (May 1973).

[2] Alam, K., K. M. L. Saxena and Y. L. Tong, "Optimal Confidence Interval for a Ranked Parameter," Journal of the American Statistical Association, *68*, 720–725 (September 1973).

[3] Chen, H. J. and E. J. Dudewicz, "Procedures for Fixed-Width Interval Estimation of the Largest Normal mean," Journal of American Statistical Association, *71*, 752–756 (September 1976).

[4] Dudewicz, E. J., "Confidence Intervals for Ranked Means," Naval Research Logistics Quarterly, *17*, 69–78 (March 1970).

[5] Dudewicz, E. J., "Two-Sided Confidence Intervals for Ranked Means," Journal of the American Statistical Association, *67*, 462–464 (June 1972).

[6] Dudewicz, E. J., "Point Estimation of Ordered Parameters: The General Location Parameter Case," Tamkang Journal of Mathematics, *3*, 101–114 (November 1972).

[7] Dudewicz, E. J., J. S. Ramberg and H. J. Chen, "New Tables for Multiple Comparisons with a Control (Unknown Variances)," Biometrische Zeitschrift, *17*, 13–26 (1975).

[8] Dudewicz, E. J. and Y. L. Tong, "Optimal Confidence Intervals for the Largest Location Parameter," *Statistical Decision Theory and Related Topics*, ed. S. S. Gupta and J. Yackel, 363–375 (Academic Press, Inc., New York, 1971).

[9] Lehmann, E. L., *Testing Statistical Hypothesis* (John Wiley and Sons, Inc., New York, 1959).

[10] Parzen, E., *Modern Probability Theory and Its Applications* (John Wiley and Sons, Inc., New York, 1960; Fourth Printing, March 1963).

[11] Pearson, K., *Tables of the Incomplete* Γ *Function*, Computed by the Staff of the Department of Applied Statistics, University of London (London, 1957).

[12] Pearson, K., *Tables of the Incomplete Beta Function* (Cambridge University Press, London, 2nd ed., 1968).

[13] Rizvi, M. H. and K. M. L. Saxena, "On Interval Estimation and Simultaneous Selection of Ordered Location and Scale Parameters," Annals of Statistics, *2*, 1340–1345 (November 1974).

[14] Saxena, K. M. L., "Interval Estimation of the Largest Variance of k Normal Populations," Journal of the American Statistical Association, *66*, 408–410 (June 1971).

[15] Saxena, K. M. L. and Y. L. Tong, "Interval Estimation of the Largest Mean of k Normal Populations with known Variance," Journal of the American Statistical Association, *64*, 296–299 (March 1969).

[16] Saxena, K. M. L. and Y. L. Tong, "Optimum Interval Estimation for the Largest Scale Parameter," Abstract, *Optimizing Methods in Statistics*, ed. J. S. Rustagi, 477 (Academic Press, Inc., New York and London, 1971).

AN $M/M/1$ QUEUE WITH DELAYED FEEDBACK*

Ingjaldur Hannibalsson

Ohio State University
Columbus Ohio

Ralph L. Disney

University of Michigan
Ann Arbor, Michigan

ABSTRACT

We present some results for $M/M/1$ queues with finite capacities with delayed feedback. The delay in the feedback to an $M/M/1$ queue is modelled as another M-server queue with a finite capacity. The steady state probabilities for the two dimensional Markov process $\{N(t), M(t)\}$ are solved when $N(t)$ = queue length at server 1 at t and $M(t)$ = queue length at server 2 at t. It is shown that a matrix operation can be performed to obtain the steady state probabilities. The eigenvalues of the operator and its eigenvectors are found. The problem is solved by fitting boundary conditions to the general solution and by normalizing. A sample problem is run to show that the solution methods can be programmed and meaningful results obtained numerically.

1. STATEMENT OF THE PROBLEMS

The problem that will be examined in this paper is one in the class of networks of queues with delayed feedback. We modeled the system as follows. This system has two servers. It is assumed that customers arrive at server I according to a Poisson process with rate λ. Server I has a waiting space for N customers, where N will be assumed to be finite. Service is on a first come, first served basis. The service times are exponentially distributed with mean $1/\mu_1$. When a customer has completed service in server I, with probability p he goes to server II and with probability $q=1-p$ he leaves the system. Server II has a waiting space for M customers, and M will be assumed to be finite. The service times at server II are assumed to be exponentially distributed with mean $1/\mu_2$. The service is on a first come, first served basis. When a customer has completed service at server II he goes back to server I with probability 1. This system will be analyzed when blocking occurs (i.e. when a customer that is supposed to enter the other server occupies his own server until a space at the other server gets empty). Only steady state probabilities are obtained.

The second server provides the delay in feedback. While such servers physically exist in some systems, our interest in them is only to have them serve as one means of providing a delay in feeding back. The delay times then are random and correspond to the time the job spends at the second server. If the blocking time at the second server is neglected in determining the feedback time, then the amount of delay is simply the total time spent at the second server.

*This research was supported in part by the United States Office of Naval Research under contract number N00014-75-C-0492 (NR 042-296). Reproduction in whole or in part is permitted for any purpose of the United States Government.

The case in which $M, N=\infty$ has been studied by Jackson [3] who found that the system acted as two independent $M/M/1$ queues. Independence is lost, however, in the above stated problem. Other work that we can find that might include the stated problem as a special case (e.g. [1, 5], likewise treat the case of $M=N=\infty$ or do not provide for blocking [4]. Since the blocking phenomenon itself is of some practical importance, we have chosen to keep M, N finite. For $M\to\infty$, the problem can be solved using a limiting argument on the results to follow when $N<\infty$. For $N\to\infty$, $M<\infty$, the following methods involve infinite dimensional matrices. Whether the methods of this paper go through in that case is presently unknown. The case where blocking does not occur and therefore some jobs do not gain access to the waiting lines, i.e. overflow, has been studied. Since the analysis is closely parallel to that given in the following study we omit it here. Full results are available in [2].

2. FORMAL STATEMENT OF THE PROBLEM AND DEFINITIONS

The queue length problem is concerned with a two dimensional, irreducible Markov chain $(N(t), M(t))$ with state space $((n, m): 0\leq n\leq N<\infty$, and $0\leq m\leq M<\infty)$. A state (n, m) of the system denotes n customers at server I and m customers at server II. N and M will be assumed to be finite. The probability of being in state (n, m) is denoted by $p_{n,m}$. Both servers are empty if $(n, m)=(0, 0)$.

It assumed that a customer, who has completed service in server I and is supposed to feed back, blocks server I if he cannot enter the waiting space of server II because it already contains M customers. It is also assumed that a customer who has completed service in server II blocks server II if he cannot enter the waiting space of server I because it already contains N customers. Arriving customers who find N present in the waiting space of server I are cleared.

$(N(t), M(t))$ is a finite, irreducible Markov chain. Λ is the infinitesimal generator of the process. In this case special blocking states have to be defined. The state when there are i customers at server I, one of which is waiting for an empty space at server II, is denoted by $((i-1, M+1): i=1, \ldots, N)$. The state, when there are j customers at server II, one of which is waiting for an empty space at server I, is denoted by $((N+1, j-1): j=1, \ldots, M)$.

By lexicographically ordering the states, the Λ matrix, which is $((N+1)(M+1)+N+M)\times((N+1)(M+1)+N+M)$ dimensional can be written as follows.

$$\Lambda=\begin{bmatrix} -A_0 & A_3 & & & & & & & \\ A_1 & -A_2 & A_3 & & & & & & \\ & \ddots & \ddots & \ddots & & & & & \\ & & \ddots & \ddots & \ddots & & & & \\ & & & \ddots & \ddots & \ddots & & & \\ & & & & A_1 & -A_2 & A_3 & & \\ & & & & & A_1 & -A_4 & A_9 & \\ & & & & & & A_7 & -A_5 & A_{10} \\ & & & & & & & A_8 & -A_6 \end{bmatrix}$$

A_0 is an $(M+2)\times(M+2)$ diagonal maxtrix with λ for its first diagonal element and $\lambda+\mu_2$ for all other diagonal elements.

A_1 is an $(M+2)\times(M+2)$ bidiagonal matrix with $q\mu_1$ everywhere on the main diagonal except that the last term is 0. $p\mu_1$ lies everywhere on the first superdiagonal.

A_2 is an $(M+2)\times(M+2)$ diagonal matrix with $\lambda+\mu_1$ for its first diagonal term, $\lambda+\mu_2$ for its last diagonal term and $\lambda+\mu_1+\mu_2$ everywhere else on the diagonal.

A_3 is an $(M+2)\times(M+2)$ bidiagonal matrix with λ everywhere on the main diagonal and μ_2 on the first subdiagonal.

A_4 is an $(M+2)\times(M+2)$ diagonal matrix with $\lambda+\mu_1$ for its first element, $\lambda+\mu_1+\mu_2$ for the remaining diagonal elements except the last which is μ_2.

A_5 is an $(M+1)\times(M+1)$ diagonal matrix with μ_1 for its first element and $\mu_1+\mu_2$ elsewhere on the diagonal.

A_6 is an $M\times M$ diagonal matrix with μ_1 on the diagonal.

A_7 is $(M+1)\times(M+2)$ bidiagonal matrix with $q\mu_1$ on the main diagonal and $p\mu_1$ on the first superdiagonal.

A_8 is an $M\times(M+1)$ bidiagonal matrix with the same diagonal elements as A_7.

A_9 is an $(M+2)\times(M+1)$ bidiagonal matrix with λ on the main diagonal and μ_2 on the first subdiagonal.

A_{10} is an $(M+1)\times M$ matrix with μ_2 everywhere on the first subdiagonal and zero everywhere else.

As the Markov process is finite and irreducible, there exists a unique probability vector $\underset{\sim}{P}$ with its elements ordered lexicographically, satisfying

$$\text{(2.1a)}\ \underset{\sim}{P}\Lambda=0, \qquad \text{(2.1b)}\ 0<\underset{\sim}{P}<1, \qquad \text{(2.1c)}\ \sum_n \sum_m p_{nm}=1.$$

The elements of $\underset{\sim}{P}$ are the usual steady state probabilities for a finite dimensional Markov chain.

Letting $\underset{\sim}{P}=(P_0, P_1, \ldots, P_N, P_{N+1})$ where $P_k=(p_{k,0}, p_{k,1}, \ldots, p_{k,M}, p_{k,M+1})$, for $k=0, 1, \ldots, N-1$, $P_N=(p_{N,0}, p_{N,1}, \ldots, p_{N,M})$ and $P_{N+1}=(p_{N+1,0}, p_{N+1,1}, \ldots, p_{N+1,M-1})$, (2.1.a) can be rewritten as

$$\text{(2.2.a)} \qquad P_N A_{10}-P_{N+1}A_6=0,$$

$$\text{(2.2.b)} \qquad P_{N-1}A_9-P_N A_5+P_{N+1}A_8=0,$$

$$\text{(2.2.c)} \qquad P_{N-2}A_3-P_{N-1}A_4+P_N A_7=0,$$

$$\text{(2.2.d)} \qquad P_k A_3-P_{k+1}A_2+P_{k+2}A_1=0, \qquad k=N-3, \ldots, 1, 0,$$

$$\text{(2.2.e)} \qquad -P_0 A_0+P_1 A_1=0.$$

If (2.2) is augmented by

$$\text{(2.3)} \qquad P_{-1}A_3=P_0(A_2-A_0)=P_0 H$$

and (2.3) is added to (2.2.e), the following system of equations results:

$$\text{(2.4.a)} \qquad P_N A_{10}-P_{N+1}A_6=0,$$

(2.4.b) $$P_{N-1}A_9-P_NA_5+P_{N+1}A_8=0,$$

(2.4.c) $$P_{N-2}A_3-P_{N-1}A_4+P_NA_7=0,$$

(2.4.d) $$P_kA_3-P_{k+1}A_2+P_{k+2}A_1=0, \qquad k=N-3, \ldots, 1, 0, -1,$$

(2.4.e) $$P_{-1}A_3=P_0H.$$

When (2.4.c), (2.4.d), and (2.4.e) are multiplied from the right by A_3^{-1} which exists since $p, q>0$, the following system of equations results,

(2.5.a) $$P_NA_{10}-P_{N+1}A_6=0,$$

(2.5.b) $$P_{N-1}A_9-P_NA_5+P_{N+1}A_8=0,$$

(2.5.c) $$P_{N-2}-P_{N-1}C_4+P_NC_7=0,$$

(2.5.d) $$P_k-P_{k+1}C_2+P_{k+2}C_1=0, \qquad k=N-3, \ldots, 1, 0, -1,$$

(2.5.e) $$P_{-1}=P_0H_C.$$

Where $C_1=A_1A_3^{-1}$, $C_2=A_2A_3^{-1}$, $C_4=A_4A_3^{-1}$, $C_7=A_7A_3^{-1}$ and $H_C=HA_3^{-1}$.

3. THE GENERAL SOLUTION

Let $X_n=(P_n, P_{n-1})$, $n=0, 1, \ldots, N-1$, be a $2M+4$ dimensional row vector, $X_N=(P_N, P_{N-1})$ a $2M+3$ dimensional vector and $X_{N+1}=(P_{N+1}, P_N)$ a $2M+2$ dimensional vector. Define a $(2M+4)\times(2M+4)$ matrix B, where

$$B=\begin{bmatrix} 0 & -C_1 \\ I & C_2 \end{bmatrix}$$

It follows from equation (2.5.d) that

(3.1) $$X_{n-1}=X_nB, \qquad n=N-1, \ldots, 2, 1.$$

Using simple iteration on (3.1)

(3.2) $$X_n=X_{N-1}B^{N-n-1}, \qquad n=N-1, \ldots, 1, 0.$$

Thus (3.2) determines the vector X_n, $n=N-1, \ldots, 1, 0$, in terms of the elements of X_{N-1}. Equations (2.5.a), (2.5.b) and (2.5.c) relate P_{N-2}, P_{N-1}, P_{N+1}. When (2.5.a) is written out explicitly one gets

(3.3) $$\mu_2p_{N,i}=\mu_1p_{N+1,i-1}, \qquad i=1, 2, \ldots, M.$$

Thus the component of P_N can be written in terms of the components of P_{N+1} and $p_{N,0}$. When (2.5.b) is written out explicitly one gets

$$p_{N-1,0}+\mu_2p_{N-1,1}-\mu_1p_{N,0}+q\mu_1p_{N+1,0}=0,$$

$$p_{N-1,i}+\mu_2p_{N-1,i+1}-(\mu_1+\mu_2)p_{N,i}+p\mu_1p_{N+1,i-1}+q\mu_1p_{N+1,i}=0, \qquad i=1, 2, \ldots, M-1,$$

$$p_{N-1,M}+\mu_2p_{N-1,M+1}-(\mu_1+\mu_2)p_{N,M}+p\mu_1p_{N+1,M-1}=0.$$

By using (3.3) it is possible to solve this system of equations in terms of the components of P_{N+1}, $p_{N,0}$ and $p_{N-1,0}$. Then by using (2.5.c) one obtains P_{N-2} in terms of the components of P_{N+1}, $p_{N,0}$ and

$p_{N-1,0}$. Thus X_n can be written uniquely in terms of the components of P_{N+1}, $p_{N,0}$ and $p_{N-1,0}$, for $n=0, 1, \ldots, N+1$. Therefore P_n can be written uniquely in terms of the components of P_{N+1}, $p_{N,0}$ and $p_{n-1,0}$, and by using (2.1.c) and (2.5.e) P_n can be determined uniquely for $n=0, 1, \ldots, N+1$.

The major problem is to determined B^{N-n-1} explicitly. A similarity transformation on B as $B=RJR^{-1}$ is sought. In Section 4 it well be shown that B has $2M+4$ distinct eigenvalues two of which are 0 and 1. Thus J is diagonal. In Section 5 it is shown that both R and R^{-1} can be written out explicity, and finally in Section 6 the boundary conditions are fitted.

4. THE EIGENVALUE PROBLEM

In this section the eigenvalues of B well be found.

LEMMA 4.1: The invariant polynomials of B that are not identically 1 are those of the z-matrix $z^2I-zC_2+C_3$.

PROOF: The characteristic matrix for B is

$$B(z)=(zI-B)=\begin{bmatrix} zI & C_1 \\ -I & zI-C_2 \end{bmatrix}.$$

Let

$$S(z)=\begin{bmatrix} 0 & -I \\ I & zI \end{bmatrix} \text{ and } T(z)=\begin{bmatrix} I & zI-C_2 \\ 0 & I \end{bmatrix}$$

$S(z)$ and $T(z)$ were determined in order to have

$$S(z)B(z)T(z)=\begin{bmatrix} I & 0 \\ 0 & z^2I-zC_2+C_1 \end{bmatrix}.$$

As $S(z)B(z)T(z)$ is an equivalence transformation of $B(z)$ and since equivalent matrices have the same invariant polynomials, the invariant polynomials of $B(z)$, that are not identically 1 are among those of the matrix $z^2I-zC_2+C_1$. By performing row and column operations, $z^2I-zC_2+C_1$ can be written in the following form.

$$z^2I-zC_2+C_1=\begin{bmatrix} a & e & & & & \\ d & b & e & & & \\ & d & b & e & & \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & \ddots & \\ & & & & \ddots & \\ & & & d & b & e \\ & & & & d & c \end{bmatrix}$$

which is $(M+2)\times(M+2)$ dimensional

$$a=z^2-(\lambda+\mu_1)z/\lambda+\mu_1 q/\lambda,$$

$$b=z^2-(\lambda+\mu_1+\mu_2)z/\lambda+\mu_1 q/\lambda,$$

$$c=z^2-(\lambda+\mu_2)z/\lambda,$$

$$d=\mu_2 z^2/\lambda,$$

$$e=p\mu_1/\lambda.$$

Let D_n, and $n\times n$ determinant, be defined recursively by

$$D_1=z(z-1),$$

$$D_2=z(z-1)(z^2-(\lambda+\mu_1+\mu_2)z/\lambda+q_{\mu_1}/\lambda+q_{\mu_1\mu_2}z^2/\lambda^2),$$

$$D_n=(z^2-(\lambda+\mu_1+\mu_2)z/\lambda+q_{\mu_1}/\lambda)D_{n-1}-(p_{\mu_1\mu_2}z^2/\lambda^2)D_{n-2}, \qquad n=3, 4, \ldots, M+2.$$

LEMMA 4.2:

$$D_{M+2}=|z^2I-zC_2+C_1|.$$

PROOF: This is proved by performing elementary row and column operations on D_{M+2}.

K_n is defined to be $K_n(z)=D_n/z(z-1)$.

The z argument will be omitted when not necessary. Clearly;

$$K_n=(z^2-(\lambda+\mu_1+\mu_2)z/\lambda+q_{\mu_1}/\lambda)K_{n-1}-(p\mu_1\mu_2 z^2/\lambda^2)K_{n-2}, \qquad n=3, 4, \ldots, M+2,$$

and

$$K_2=(z^2-(\lambda+\mu_1+\mu_2)z/\lambda+q\mu_1\mu_2/\lambda^2),$$

(4.1)

$$K_1=1.$$

The following values of K_n (z) will be useful later. We list them here for reference and call them formula (4.2).

(a) $K_n(1)=((-1)^{n+1}(p\mu_1/\lambda)^n/(\mu_2-p\mu_1))((\mu_2-p\lambda/q)+(\lambda p/q-\mu_1 p)(\mu_2/p\mu_1)^n)$, $n=1, 2, \ldots, M+2$, when $p\mu_1\neq\mu_2$.

(4.2) (b) $K_n(1)=(-1)^{n-1}(p\mu_1/\lambda)^{n-1}(q\mu_1/\lambda+n(1-q\mu_1/\lambda))$, $n=1, 2, \ldots, M+2$ when $p\mu_1=\mu_2$.

(c) $K_n(q\mu_1/\lambda)=(-1)^{n-1}(qp\mu_1^2/\lambda^2)^{n-1}$, $n=1, 2, \ldots, M+2$.

In order to show, that K_{M+2} has $2M+2$ distinct roots, which implies that D_{M+2} has $2M+4$ distinct roots, it has to be shown how K_n (1) changes sign as a function of n. Due to the complexity of K_n (1) several cases have to be analyzed. Let

$$h_n=|(p\mu_1/\lambda)^n(p/q)/(\mu_2-p\mu_1)|,\ a=(\mu_2-\lambda p/q),\ b=(\lambda p/q-\mu_1 p)$$

and

$$c=(\mu_2/p\mu_1),\ d=q\mu_1/\lambda) \text{ and } e=(1-q\mu_1/\lambda).$$

The cases, when $p\mu_1 \neq \mu_2$ are shown in Table 1.

TABLE 1. The form of $K_n(1)$, when $p\mu_1 \neq \mu_2$

case	The form of $K_n(1)$	a	b	c
1	$(-1)^{n+1}h_n(+\|a\|+\|b\|c^n)$	>0	>0	>1
2	$(-1)^{n+1}h_n(+\|a\|-\|b\|c^n)$	>0	<0	>1
3	$(-1)^{n+1}h_n(-\|a\|+\|b\|c^n)$	<0	>0	>1
4	$(-1)^{n+1}h_n(-\|a\|-\|b\|c^n)$	<0	<0	>1
5	$(-1)^{n+2}h_n(+\|a\|+\|b\|c^n)$	>0	>0	<1
6	$(-1)^{n+2}h_n(+\|a\|-\|b\|c^n)$	>0	<0	<1
7	$(-1)^{n+2}h_n(-\|a\|+\|b\|c^n)$	<0	>0	<1
8	$(-1)^{n+2}h_n(-\|a\|-\|b\|c^n)$	<0	<0	<1
9	$(-1)^{n+1}h_n\|b\|c^n$	$=0$	>0	>1
10	$(-1)^{n+2}h_n\|b\|c^n$	$=0$	<0	<1
11	$(-1)^{n+1}h_n\|a\|$	>0	$=0$	>1
12	$(-1)^{n+2}h_n\|a\|$	<0	$=0$	<1

When $p\mu_1-\mu_2=a+b>0$, $c>1$. Thus cases 4 and 5 are not possible. The cases when $p\mu_1=\mu_2$ are shown in Table 2. Let $d=q\mu_1/\lambda$ and $e=(1-q\mu_1/\lambda)$.

TABLE 2. The form of K_n (1), when $p\mu_1=\mu_2$.

case	The form of $K_n(1)$	d	e
1	$(-1)^{n+1}(p\mu_1/\lambda)^{n-1}(+d+ne)$	>0	>0
2	$(-1)^{n+1}(p\mu_1/\lambda)^{n-1}(+d-ne)$	>0	<0
3	$(-1)^{n+1}(p\mu_1/\lambda)^{n-1}(+d)$	>0	$=0$

As the analyses of all cases are similar, the analysis for case 1, when $p\mu_1 \neq \mu_2$ will be shown. Results for all the others follow a similar pattern. As $b=\lambda p/q-\mu_1 p>0$, it is clear, that $q\mu_1/\lambda<1$.

LEMMA 4.3: $K_2(z)$ has two roots α_1 and α_2, one of which lies in $[0, q\mu_1/\lambda]$ and the other in $[1, \infty]$.

PROOF: The proof is easy since $K_2(z)$ is a quadratic function of z.

LEMMA: 4.4: $K_3(z)$ has four roots. Two lie in $[0, q\mu_1/\lambda]$ and two lie in $[1, \infty]$. The four roots are distinct.

PROOF: By the definition of $K_3(z)$, formula (4.1), it is clear that $K_3(0)>0$ and $K_3(\alpha_1)<0$. From property (4.2.c), $K_3(q\mu_1/\lambda)>0$. Thus $K_3(z)$ has at least two roots in $[0, q\mu_1/\lambda]$ and α_1 separates two of them. From property (4.2.a) and the definition of $K_3(z)$, $K_3(1)>0$, $K_3(\alpha_2)<0$. Furthermore the definition of $K_3(z)$, implies $K_3(z)\to\infty$ for $z\to\infty$. Thus $K_3(z)$ has at least two roots in $[1, \infty)$ and α_2 separates two of them. Since $K_3(z)$ has only four roots, two lie in $[0, q\mu_1/\lambda]$ and two lie in $[1, \infty)$ and the four roots are distinct.

THEOREM 4.5: $K_n(z)$ has $2n-2$ distinct roots $\gamma_1, \gamma_2, \ldots, \gamma_{2n-2}$, $n-1$ of which is in $[0, q\mu_1/\lambda]$ and $n-1$ of which are in $[1, \infty)$. If $\eta_1, \eta_2, \ldots, \eta_{2p-4}$ are the roots of $K_{p-1}(z)$ and $\phi_1, \phi_2, \ldots, \phi_{2p-4}, \phi_{2p-3}, \phi_{2p-2}$ are the roots of $K_p(z)$, one has

$$0<\phi_1<\eta_1<\phi_2< \cdots <\eta_{p-2}<\phi_{p-1}<q\mu_1/\lambda<1<\phi_p<\eta_{p-1}< \cdots <\eta_{2p-4}<\phi_{2p-2}<\infty$$

PROOF: This theorem is proved by induction from Lemma 4.4.

Using the same technique it is easy to show for all the other cases that K_{M+2} has $2M+2$ distinct roots, γ_i, and

$$J=\begin{bmatrix} 0 & & & & \\ & 1 & & & \\ & & \gamma_1 & & \\ & & & \ddots & \\ & & & & \gamma_{2M+1} \\ & & & & \\ & & & & \gamma_{2M+2} \end{bmatrix}$$

5. THE EIGENVECTOR PROBLEM

To find the right hand eigenvectors of B one looks at

$$(B-\gamma I)x=0. \tag{5.1}$$

and

$$y(B-\gamma I)=0. \tag{5.2}$$

First the right hand eigenvectors will be found. By performing row operations on $(B-\gamma I)$ the system of equations. (5.1) reduces to the following system to equations.

$$x_0+(c-\gamma)x_{M+2}=0, \tag{5.3}$$

$$x_1-(\mu_2/\lambda)(d/c)\gamma x_{M+2}+(d-\gamma)x_{M+3}+(\mu_2/\lambda)(d/c)x_0=0, \tag{5.4}$$

$$x_i-(\mu_2/\lambda)\gamma x_{M+i+1}+(d-\gamma)x_{M+i+2}+(\mu_2/\lambda)x_{i-1}=0, \qquad i=2, 3, \ldots, M, \tag{5.5}$$

$$x_{M+1}-(\mu_2/\lambda)(e/d)\gamma x_{2M+2}+(e-\gamma)x_{2M+3}+(\mu_2/\lambda)(e/d)x_M=0, \tag{5.6}$$

$$(a/c)\gamma x_{M+2}+bx_{M+3}-(a/c-\gamma)x_0=0, \tag{5.7}$$

$$(a/d)\gamma x_{M+i}+bx_{M+i+1}-(a/d-\gamma)x_{i-2}=0, \qquad i=3, \ldots, M+2, \tag{5.8}$$

$$x_{M+1}=0, \tag{5.9}$$

where

$$a=(\mu_1/\lambda)(q-p\mu_2/\lambda), \qquad d=(\lambda+\mu_1+\mu_2)/\lambda,$$
$$b=p\mu_1/\lambda, \qquad e=(\lambda+\mu_2)/\lambda.$$
$$c=(\lambda+\mu_1)/\lambda,$$

The solution can be obtained using standard difference equation techniques by using equation 5.3–5.5 and 5.7–5.9. Then 5.6 is satisfied if and only if γ is an eigenvalue.

To find the left hand eigenvectors column operations are performed on $(B-\gamma I)$. Then the system of equations reduces to the following system of equations.

$$\gamma y_i = y_{M+i+2}, \qquad i=0, 1, \ldots, M+1, \tag{5.10}$$

$$(c-\gamma)y_{M+2}-(\mu_2/\lambda)\gamma y_{M+3}-(a+\mu_2 b/\lambda)y_0=0, \tag{5.11}$$

$$(d-\gamma)y_{M+i+2}-(\mu_2/\lambda)y_{M+i+3}-by_{i-1}-(a+\mu_2 b/\lambda)y_i=0, \qquad i=1, 2, \ldots, M, \tag{5.12}$$

$$(e-\gamma)y_{2M+3}-by_M=0. \tag{5.13}$$

where a, b, c, d and e are the same as before. The solution can be obtained using standard difference equation techniques. By using equations (5.10)–(5.12), and then (5.13) is satisfied if and only if γ is an eigenvalue.

6. BOUNDARY CONDITIONS

It has already been shown that

$$X_0=X_{N-1}B^{N-1}=X_{N-1}RJ^{N-1}R^{-1}, \tag{6.1}$$

when the definition of X_i is used, (6.1) turns out to be equivalent to

$$(P_0,P_{-1})=(P_{N-1},P_{N-2})\,RJ^{N-1}R^{-1}. \tag{6.2}$$

P_{N-1} and P_{N-2} can be written in terms of the components of P_{N+1}, $p_{N,0}$ and $p_{N-1,0}$. (6.2) contains $2M+4$ scalar equations. The variables in those equations are $p_{N+1,0}, \ldots, p_{N+1,M-1}, p_{N,0}, p_{N-1,0}, p_{0,0}, \ldots, p_{0,M+1}$ and $p_{-1,0}, \ldots, p_{-1,M+1}$. (6.2) can be solved for $p_{0,0}, \ldots, p_{0,M+1}$ and $p_{-1,0}, \ldots, p_{-1,M+1}$ in terms of $p_{N+1,0}, \ldots, p_{N+1,M-1}, p_{N,0}$ and $p_{N-1,0}$. From (2.5.e) one obtains $p_{-1,0}, \ldots, p_{-1,M+1}$ as linear functions of $p_{0,0}, \ldots, p_{0,M+1}$. When the results of (6.2) are substituted into (2.5.e) one obtains $M+2$ linear equations, $M+1$ of which are independent. These equations determine $p_{N+1,0}, \ldots, p_{N+1,M-1}$, $p_{N,0}$ and $p_{N-1,0}$ within a multiplicative constant. Then by using (2.1.c)

$$\sum p_{ij}=1$$

the normalization constant can be determined. Thus the probability vector $\underset{\sim}{P}$ has been determined uniquely.

THEOREM 6.1: The finite feedback problem with blocking, defined by (2.1), (2.5) and the corresponding matrices defined in section 2, has a unique probability solution (X_n)

$$X_n=X_{N-1}B^{N-n-1}, \qquad n=N-1, \ldots, 1, 0,$$

where $B=RJR^{-1}$ whose terms have been determined in section 5. Furthermore $X_0=(P_0, P_{-1})$ satisfies (6.2) which together with (2, 5. e) and (2, 1. c) determines X_{N-1} uniquely.

PROOF: The proof is included in Section 4, 5, and 6.

7. AN EXAMPLE

In this section an example is given. It is assumed that $N=3$ and $M=1$, $\lambda=1$, $\mu_1=4$, $\mu_2=2$ and $p=q=1/2$. The B matrix is found to be

$$B=\begin{bmatrix} 0 & 0 & 0 & 2 & -2 & 0 \\ 0 & 0 & 0 & -4 & 2 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 5 & 0 & 0 \\ 0 & 1 & 0 & -14 & 7 & 0 \\ 0 & 0 & 1 & 12 & -6 & 3 \end{bmatrix}$$

Then it follows that

$$J=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & .2859 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.2548 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3.9137 & 0 \\ 0 & 0 & 0 & 0 & 0 & 8.5456 \end{bmatrix}$$

The R matrix is found to be

$$R=\begin{bmatrix} 1.6667 & -1.2755 & .4141 & -4.2612 & .0729 & .1590 \\ -2.3333 & .7460 & .4141 & -5.9675 & -.2921 & -.3529 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ -.3333 & .2706 & -.1035 & 1.1378 & -.0671 & .0448 \\ -.3333 & .4530 & -.3106 & 3.8112 & -.2098 & -.6345 \\ .3333 & -.1950 & -.3106 & 5.2796 & .4961 & .7835 \end{bmatrix}$$

and the R^{-1} matrix

$$R^{-1}=\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ -.2427 & .9678 & 2.4940 & -.0694 & .2767 & .7131 \\ 9.6589 & 9.6589 & 9.6589 & 9.6589 & 9.6589 & 9.6589 \\ .7478 & .6411 & .5855 & .9384 & .8044 & .7347 \\ -2.8462 & -.2092 & .1170 & -11.1390 & -.8187 & .4579 \\ .5314 & .1175 & .0050 & 4.5414 & -1.0043 & .0424 \end{bmatrix}$$

Within numerical accuracy the probability distribution is found to be:

TABLE 3. The steady state probabilities when $N=3$ and $M=1$.

P_{ij}	j 0	1	2	P_{i-}
0	0. 2851	0. 1430	0. 0448	0. 4729
1	0. 1444	0. 0677	0. 0450	0. 2571
i 2	0. 0755	0. 0451	0. 0525	0. 1731
3	0. 0489	0. 0300		0. 0789
4	0. 0150			0. 0150
P_{-j}	0. 5689	0. 2858	1423	

From Table 3 one obtains the probability that server *I* is blocked as 0.1423 and the probability that server *II* is blocked as 0.0150.

ACKNOWLEDGMENTS

This work was completed while the second author was Distinguished Visiting Professor at the Ohio State University. He would like to thank the faculty, staff and students of that university for their help in preparing this paper.

BIBLOGRAPHY

[1] Arya, K. L., "Systems of Two Servers in Bi-Series with a Serial Service Channel and Phase Type Service," Zeitschrift für Operations Research, *B4*, 115 (1970).

[2] Hannibalsson, I., "Networks of Queues with Delayed Feedback," Technical Report 75–10, Department of Industrial and Operations Engineering, University of Michigan (June 1975).

[3] Jackson, J. R., "Networks of Waiting Lines," Operations Research, *5*, 518 (1957).

[4] Jackson, J. R., "Job-Shop Like Queueing Systems," Management Science, *10*, 131 (1963).

[5] Maggu, P. L., "Phase Type Service Systems with Two Servers in Bi-Series," Journal of Operations Research Society of Japan, *42*, 505 (1962).

OPTIMAL DYNAMIC RULES FOR ASSIGNING CUSTOMERS TO SERVERS IN A HETEROGENEOUS QUEUING SYSTEM

Wayne Winston

Indiana University
Bloomington, Indiana

ABSTRACT

We consider a queuing system in which both customers and servers may be of several types. The distribution of a customer's service time is assumed to depend on both the customer's type and the type of server to which he is assigned. For a model with two servers and two customer types, conditions are presented which ensure that the discounted number of service completions is maximized by assigning customers with longer service times to faster servers. Generalizations to more complex models are discussed.

1. INTRODUCTION

For reasons of analytic simplicity, most modelers of congested systems assume that all servers are identical. There are many situations in which such an assumption is unrealistic. For example, in a hospital the assumption of homogeneous servers equates a semi-private room with a coronary care unit. In a supermarket, the assumption of identical servers ignores the express lane and the different rates at which cashiers work. A queuing system in which both servers and customers are of several types will be called a *heterogeneous queuing system*. In a heterogeneous queuing system the method used to assign customers to servers is an important aspect of the system's operation. Let the state of the system be defined by knowledge of the type of customer (if any) occupying each server. Most prior work on the assignment of customers in heterogeneous congestion systems assumes that customers are assigned to servers according to rules that are independent of the state of the system. (c. f. Kotiah and Slater [6] and Rolfe [8].) In this paper the assignment rules under consideration will depend on the system state.

2. THE MODEL

Consider a queuing system consisting of s servers. Customers of type i ($i=1, 2, \ldots, r$) arrive at rates λ_i according to independent Poisson processes. Upon arrival, a customer must be assigned to an idle server; if all servers are occupied, then an arrival is lost to the system. A type i customer who is assigned to server j completes service according to an exponential distribution with parameter μ_{ij}. A unit reward is earned whenever a customer completes service. Rewards are assumed

to be discounted by a factor α, so a unit reward earned at time t is equivalent to a reward of $e^{-\alpha t}$ earned at time 0. The goal is to assign customers to servers so as to maximize the expected discounted reward earned over an infinite horizon.

As an example of a situation where the above model may be applicable let the servers be fire engines and the customers be fire alarms. Then a fire engine must be dispatched whenever an alarm is recorded. An alarm that is recorded when no fire engine is available is considered to be lost to the system because the chance of controlling a fire is assumed to be nil if an engine is not immediately dispatched. In this context, the μ_{ij}'s would depend on the location of the alarms and fire engines. For further work on the problem of dispatching fire engines the reader is referred to [1].

It is clear that the problem of determining an optimal assignment policy may be formulated as a continuous time Markov decision process (CTMDP); see Howard [4]. To do so define $I_r=\{0, 1, \ldots, r\}$ and let $(I_r)^s$ be the s-fold Carterian product of I_r. Finally, for $C=(n_1, n_2, \ldots, n_s) \in (I_r)^s$ let $X(C)=\{i|n_i=0\}$. Then the relevant CTMDP is in state $C=(n_1, \ldots, n_s) \in (I_r)^s$ whenever server i is occupied by a type n_i customer ($n_i=0$ indicating that server i is unoccupied). For any state C having $X(C)\neq\{\}$ the set of possible actions is F_C, the set of all functions mapping $I_r-\{0\}\rightarrow X(C)$; for all C such that $X(C)=\{\}$ the only action is a dummy action, labeled $s+1$, which has no effect on the process. When action $s+1$ is chosen all arrivals are turned away and are lost to the system. If an action $f \in F_C$ such that $f(i)=k$ is chosen then any type i arrival who finds the system in state C will be assigned to server k.

We now assume that the μ_{ij}'s satisfy the following conditions

(1) $$\mu_{il}+\mu_{jk}\geq\mu_{ik}+\mu_{jl}, \qquad j>i,\ l>k$$

(2) $$\mu_{jk}\geq\mu_{ik}, \qquad j>i$$

and

(3) $$\mu_{il}\geq\mu_{ik} \qquad l>k$$

Condition (2) implies that lower numbered customer types have longer expected service times while condition (3) implies that higher numbered servers are faster.

Let π_α be a stationary policy which is optimal when the discount factor is α. In [9] it was conjectured that conditions (1)—(3) ensure that the action f_C chosen by π_α whenever the state is C satisfies

(4) $$f_C(i)\geq f_C(i+1) \qquad C\in(I_r)^s,\ 1\leq i\leq r-1.$$

By (2) and (3), a policy that satisfies (4) assigns customers with longer service times to faster servers; such a policy will be called a *longest in fastest* (LIF) policy. In [9] it was shown that a LIF policy is optimal for a two period discrete time version of the above model. We content ourselves with proving the optimality of LIF for the case $r=s=2$.

3. OPTIMALITY OF LIF FOR A TWO SERVER MODEL

If we write $A=(0, 0)$, $B=(0, 1)$, $C=(0, 2)$, $D=(1, 0)$, $E=(2, 0)$, $F=(1, 2)$, $G=(1, 1)$, $H=(2, 1)$, and $I=(2, 2)$, then for $r=s=2$ the CTMDP associated with $r=s=2$ (herefter abbreviated as CT)

has a state space $S=\{A, B, C, D, E, F, G, H, I\}$. We let f_{sk} denote the k'th possible action in state s. Then the set of possible actions may be written as

Actions

$$f_{A1}(1)=2,\ f_{A1}(2)=2,$$

$$f_{A2}(1)=1,\ f_{A2}(2)=1,$$

$$f_{A3}(1)=1,\ f_{A3}(2)=2,$$

$$f_{A4}(1)=2,\ f_{A4}(2)=1,$$

$$f_{B1}(1)=f_{B1}(2)=f_{C1}(1)=f_{C2}(2)=1,$$

$$f_{D1}(1)=f_{D1}(2)=f_{E1}(1)=f_{E1}(2)=2,$$

$$f_{F1}(1)=f_{F1}(2)=f_{G1}(1)=f_{G1}(2)=f_{H1}(1)=f_{H1}(2)$$
$$=f_{I1}(1)=f_{I2}(1)=3.$$

When action k is chosen in state i the transition rate from i to j will be written as a_{ij}^{k}. The transition rates for CT may then be written as

$$a_{AB}^{1}=\lambda_1,\ a_{AC}^{1}=\lambda_2,\ a_{AD}^{2}=\lambda_1,\ a_{AE}^{2}=\lambda_2,\ a_{AC}^{3}=\lambda_2,\ a_{AD}^{3}=\lambda_1,$$

$$a_{AB}^{4}=\lambda_1,\ a_{AE}^{4}=\lambda_2,\ a_{BA}^{1}=\mu_{12},\ a_{BG}^{1}=\lambda_1,\ a_{BH}^{1}=\lambda_2,\ a_{CA}^{1}=\mu_{22},$$

$$a_{CF}^{1}=\lambda_1,\ a_{CI}^{1}=\lambda_2,\ a_{DA}^{1}=\mu_{11},\ a_{DF}^{1}=\lambda_2,\ a_{DG}^{1}=\lambda_1,\ a_{EA}^{1}=\mu_{21},$$

$$a_{EH}^{1}=\lambda_1,\ a_{EI}^{1}=\lambda_2,\ a_{FC}^{1}=\mu_{11},\ a_{FD}^{1}=\mu_{22},\ a_{GB}^{1}=\mu_{11},\ a_{GD}^{1}=\mu_{12},$$

$$a_{HB}^{1}=\mu_{21},\ a_{HE}^{1}=\mu_{12},\ a_{IC}^{1}=\mu_{21},\ a_{IE}^{1}=\mu_{22}.$$

Defining q_i^k to be the rate at which rewards are earned when the state is i and action k is chosen we have

$$q_A^{1}=q_A^{2}=q_A^{3}=q_A^{4}=0,\ q_B^{1}=\mu_{12},\ q_C^{1}=\mu_{22},\ q_D^{1}=\mu_{11},\ q_E^{1}=\mu_{21},$$

$$q_F^{1}=\mu_{11}+\mu_{22},\ q_G^{1}=\mu_{11}+\mu_{12},\ q_H^{1}=\mu_{21}+\mu_{12},\ q_I^{1}=\mu_{21}+\mu_{22}.$$

We now assume that (1)—(3) are valid for the case $r=s=2$; that is, for $i=1, 2$,

$$\mu_{12}+\mu_{21}\geq\mu_{11}+\mu_{22}, \tag{5}$$

$$\mu_{2i}\geq\mu_{1i}, \tag{6}$$

and

$$\mu_{i2}\geq\mu_{i1}. \tag{7}$$

For the present problem the only times at which a non-trivial decision can be made are when an arrival finds both servers idle. Our goal is to prove the following result:

THEOREM 1: If (5)–(7) are valid, then for any discount factor α, the expected discounted reward earned over an infinite horizon is maximized by a policy that always chooses action 1 or 4 in state A.

As desired, this result shows that (5)–(7) ensure that an LIF policy is optimal for the case $r=s=2$.

To prove Theorem 1 we will consider a discrete time Markov decision process, DT, that is computationally equivalent to CT. By *computationally equivalent*, we mean that CT and DT have identical state and action spaces and for any discount factor α there exists a discount factor β, such that the α-optimal stationary policies for CT are identical to the β-optimal stationary policies for DT. Let CT be described by a state space S, action spaces A_i, transition rates $\{a_{ij}^k\}$ and reward rates $\{q_i^k\}$. If rewards in CT are discounted by α it follows from page 121 of [4] that for

$$N \geq \max_{\substack{i \in S \\ k \in A_i}} |a_{ij}^k|$$

CT is computationally equivalent to the discrete time Markov decision process described by the following:

(8)

State Space	S	
Action Spaces	A_i	$i \in S$
Transition Probabilities	$P_{ij}^k = \delta_{ij} + a_{ij}^k/N$,	$i, j \in S$; $k \in A_i$
where	$\delta_{ij} = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}$	
Discount Factor	$\beta = N/(N+\alpha)$	
Rewards	$r_i^k = \beta q_i^k/N$	$i \in S$, $k \in A_i$.

We note that the use of a discrete time Markov decision process to gain insight into the structure of an optimal policy for a continuous time Markov decision process has recently been used (with great success) by Lippman [7]. If we choose the N in (8) to be the Λ defined on page 690 of [7], then a discrete time Markov decision process identical to the one considered by Lippman is obtained. For our purposes we define DT to be the version of (8) associated with

$$N = \lambda_1 + \lambda_2 + \mu_{11} + \mu_{22} + \mu_{21} + \mu_{22}.$$

Then the proof of Theorem 1 is equivalent to

THEOREM 1′: If (5)–(7) hold, then action 1 or action 4 is optimal in state A of DT.

Before proving Theorem 1 we need to consider the problem of operating DT for $n < \infty$ periods so as to maximize expected discounted reward. A straight-forward inductive argument (see Theorem 2.1 of [3]) shows that, among randomized rules that depend on the past history of the system, the expected discounted reward earned during n periods is maximized by a rule that depends only on the present state of DT and the number of periods for which DT is to be operated. Let R be the class of all such rules. A typical $r \in R$ may be written as $r = (r^1, r^2, \ldots)$ where r^k: $\{A\} \rightarrow \{1, 2, 3, 4\}$. If $r^k(A) = j$, then action j is chosen if DT is in state A and is to be operated for k periods; in all other states, there is no freedom to choose an action so action 1 is always taken. Let $F_n[r|(i, j)]$ be the expected discounted reward earned in operating DT for n periods when rule r is followed and

the initial state is (i, j). The above remarks imply that there exists a $\bar{r}=(\bar{r}^1, \bar{r}^2, \ldots) \epsilon R$ with the property that for all (i, j) and $n \geq 1$

$$F_n[\bar{r}|(i, j)] \geq F_n[r|(i, j)] \qquad r \epsilon R.$$

Define

$$F_n(i, j)=\begin{cases} F_n[\bar{r}|(i, j)] & n \geq 1 \\ 0 & n=0, \end{cases}$$

$\overline{A}=(\lambda_1+\lambda_2)/N$, $A_1=\lambda_1/N$, $A_2=\lambda_2/N$ and

$S_{ij}=\mu_{ij}/N$.

Then for $i, j=1, 2$,

(9a) $$F_n(i, 0)=\beta S_{i1}+\beta S_{i1}F_{n-1}(0, 0)+\beta A_1 F_{n-1}(i, 1)+\beta A_2 F_{n-1}(i, 2)+\beta(1-\overline{A}-S_{i1})F_{n-1}(i, 0),$$

(9b) $$F_n(0, i)=\beta S_{i2}+\beta S_{i2}F_{n-1}(0, 0)+\beta A_1 F_{n-1}(1, i)+\beta A_2 F_{n-1}(2, i)+\beta(1-\overline{A}-S_{i2})F_{n-1}(0, i),$$

(9c) $$F_n(i, j)=\beta S_{i1}+\beta S_{j2}+\beta S_{i1}F_{n-1}(0, j)+\beta S_{j2}F_{n-1}(i, 0)+\beta(1-S_{i1}-S_{j2})F_{n-1}(i, j), \text{ and}$$

(9d) $$F_n(0, 0)=\max \begin{cases} \beta A_1 F_{n-1}(1, 0)+\beta A_2 \max [F_{n-1}(2, 0), F_{n-1}(0, 2)]+\beta(1-\overline{A})F_{n-1}(0, 0)=G_n(1) \\ \beta A_1 F_{n-1}(0, 1)+\beta A_2 \max [F_{n-1}(2, 0), F_{n-1}(0, 2)]+\beta(1-\overline{A})F_{n-1}(0, 0)=G_n(2) \end{cases}$$

Our development will require the following properties of $F_n(i, j)$

LEMMA 1: For $n \geq 1$, $F_n(0, 1) \geq F_n(1, 0)$.

LEMMA 2: For $n \geq 1$, $F_n(2, 1) \geq F_n(1, 2)$.

LEMMA 3: For $n \geq 1$, $F_n(0, 1)-F_n(1, 0) \geq F_n(0, 2)-F_n(2, 0)$.

LEMMA 4: For $n \geq 1$,

(10) $$1+F_n(0, 2) \geq F_n(1, 2),$$

(11) $$1+F_n(1, 0) \geq F_n(1, 1),$$

(12) $$1+F_n(0, 0) \geq F_n(0, 1),$$

(13) $$1+F_n(2, 0) \geq F_n(2, 1),$$

(14) $$1+F_n(0, 1) \geq F_n(1, 1),$$

and

(15) $$1+F_n(0, 0) \geq F_n(1, 0).$$

LEMMA 5. For $n \geq 1$,

(16) $$F_n(1, 2) \geq F_n(1, 1),$$

(17) $$F_n(0, 2) \geq F_n(0, 1),$$

and

(18) $$F_n(2, 2) \geq F_n(2, 1).$$

Lemmas 1–5 will be proven by induction. The trivial verification of Lemmas 1–5 for $n=1$ is omitted. To complete the inductive proof we assume the Lemmas are valid for $n-1$ and verify them for n.

PROOF OF LEMMA 1:

$$F_n(0,1)=\beta A_1 F_{n-1}(1,1)+\beta A_2 F_{n-1}(2,1)+S_{12}[\beta+\beta F_{n-1}(0,0)]+\beta(1-\bar{A}-S_{12})F_{n-1}(0,1) \qquad \text{[by 9b)]}$$

$$\geq \beta A_1 F_{n-1}(1,1)+\beta A_2 F_{n-1}(1,2)+S_{12}[\beta+\beta F_{n-1}(0,0)]+\beta(1-\bar{A}-S_{12})F_{n-1}(1,0)$$

(by Lemmas 1 and 2 of induction hypothesis)

$$\geq F_n(1,0). \qquad \text{(by (7), (9a), and (15))}$$

PROOF OF LEMMA 2: Substitution of (9c) shows that Lemma 2 is equivalent to

$$(19)\quad S_{12}[\beta+\beta F_{n-1}(2,0)]+S_{21}[\beta+\beta F_{n-1}(0,1)]+\beta(1-S_{12}-S_{21})F_{n-1}(2,1)$$
$$\geq S_{11}[\beta+\beta F_{n-1}(0,2)]+S_{22}[\beta+\beta F_{n-1}(1,0)]+\beta(1-S_{11}-S_{22})F_{n-1}(1,2).$$

By Lemmas 2 and 3 of the induction hypothesis, (19) is valid if

$$(20)\quad \beta(S_{11}+S_{22}-S_{12}-S_{21})F_{n-1}(1,2)+(S_{12}-S_{11})[\beta+\beta F_{n-1}(0,2)]$$
$$+(S_{12}-S_{22})[\beta+\beta F_{n-1}(1,0)]+(S_{21}-S_{12})[\beta+\beta F_{n-1}(0,1)]\geq 0.$$

By (6) and (7) plus Lemma 1 and (17) of the induction hypothesis (20) is valid if

$$\beta(S_{12}+S_{21}-S_{11}-S_{22})[1+F_{n-1}(0,2)-F_{n-1}(1,2)]\geq 0.$$

The last inequality holds by (5) and (10) of the induction hypothesis.

PROOF OF LEMMA 3: Substitution of (9a) and (9b) shows that Lemma 3 is equivalent to

$$(21)\quad \beta A_1 F_{n-1}(1,1)+\beta A_2 F_{n-1}(2,1)+S_{12}[\beta+\beta F_{n-1}(0,0)]+\beta(1-\bar{A}-S_{12})F_{n-1}(0,1)$$
$$-\{\beta A_1 F_{n-1}(1,1)+\beta A_2 F_{n-1}(1,2)+S_{11}[\beta+\beta F_{n-1}(0,0)]+\beta(1-\bar{A}-S_{11})F_{n-1}(1,0)\}$$
$$\geq \beta A_1 F_{n-1}(1,2)+\beta A_2 F_{n-1}(2,2)+S_{22}[\beta+\beta F_{n-1}(0,0)]+\beta(1-\bar{A}-S_{22})F_{n-1}(0,2)$$
$$-\{\beta A_1 F_{n-1}(2,1)+\beta A_2 F_{n-1}(2,2)+S_{21}[\beta+\beta F_{n-1}(0,0)]+\beta(1-\bar{A}-S_{21})F_{n-1}(2,0)\}.$$

By Lemmas 2 and 3 of the induction hypothesis, (21) will hold if

$$\beta(S_{22}-S_{21})F_{n-1}(2,0)+\beta(S_{22}-S_{12})F_{n-1}(0,1)+\beta(S_{11}-S_{22})F_{n-1}(1,0)$$
$$+(S_{12}+S_{21}-S_{11}-S_{22})[\beta+\beta F_{n-1}(0,0)]\geq 0.$$

The last inequality is valid by (5)—(7) plus Lemma i, Lemma 3, (15), and (17) of the induction hypothesis.

PROOF OF LEMMA 4: In the interests of brevity only the proof of (10) is given; the proofs of (11)—(15) are similar. Substitution of (9b) and (9c) into (10) shows that (10) is equivalent to

$$(22)\quad 1-\beta(1-\bar{A})+\{\beta A_1 F_{n-1}(1,2)+\beta A_2 F_{n-1}(2,2)+S_{22}[2\beta+\beta F_{n-1}(0,0)]$$
$$+\beta(1-\bar{A}-S_{22})[1+F_{n-1}(0,2)]\geq \beta\bar{A}F_{n-1}(1,2)+S_{11}[\beta+\beta F_{n-1}(0,2)]$$
$$+S_{22}[\beta+\beta F_{n-1}(1,0)]+\beta(1-\bar{A}-S_{11}-S_{22})F_{n-1}(1,2).$$

By (10), (18), and Lemma 2 of the induction hypothesis, (22) will hold if

$$1-\beta(1-\overline{A})+S_{22}[2\beta+\beta F_{n-1}(0,0)]\geq S_{22}[\beta+\beta F_{n-1}(1,0)].$$

Since $\beta\leq 1$, the last inequality holds by (15) of the induction hypothesis.

PROOF OF LEMMA 5. In the interests of brevity only the proof of (16) is given; the proofs of (17) and (18) are similar.

Substitution of (9b) and (9c) into (16) shows that (16) is equivalent to

$$(23)\quad S_{11}[\beta+\beta F_{n-1}(0,2)]+S_{22}[\beta+\beta F_{n-1}(1,0)]+\beta(1-S_{11}-S_{22})F_{n-1}(1,2)$$
$$\geq S_{11}[\beta+\beta F_{n-1}(0,1)]+S_{12}[\beta+\beta F_{n-1}(1,0)]+\beta(1-S_{11}-S_{12})F_{n-1}(1,1).$$

By (16) and (17) of the induction hypothesis, (23) will hold if

$$(S_{22}-S_{12})[\beta+\beta F_{n-1}(1,0)]+\beta(1-S_{11}-S_{22})F_{n-1}(1,1)\geq\beta(1-S_{11}-S_{12})F_{n-1}(1,1)$$

or

$$(S_{22}-S_{12})[\beta+\beta F_{n-1}(1,0)-\beta F_{n-1}(1,1)]\geq 0.$$

The last inquality holds by (6) and (11) of the induction hypothesis.

The following result will enable us to give an easy proof of Theorem 1′.

LEMMA 6: For $n\geq 1$, $\overline{r}^n(A)\epsilon\{1,4\}$

PROOF: By (9d) the result follows immediately from Lemma 1.

We now give theproof of Theorem 1′ (and therefore Theorem 1 as well).

PROOF OF THEOREM 1′: Lemma 6 and a well-known turnpike theorem (see Theorem 7.10 of [2]) imply that if DT is operated for an infinite number of periods then it is optimal to take action 1 or 4 whenever the state is A. This, of course, is the desired result.

We conclude this section by giving a heuristic interpretation of the optimality of LIF policies when conditions (5)–(7) are satisfied. Suppose (5) is satisfied with equality; that is, $\mu_{22}-\mu_{21}=\mu_{12}-\mu_{11}$. Then the optimality of a LIF policy can be interpreted as follows: given that both customer types gain equally from assignment to the faster server, it is more important to rid the system of customers with longer service times. With more customer and server types, however, the problem of proving or disproving the optimality of LIF policies becomes much more difficult. If customers can be switched between servers, however, then Johansen [5] and Winston [9] have shown the optimality of LIF policies with respect to the criterion of maximizing, with respect to stochastic order, the number of customers to complete their service in any time $t<\infty$.

4. EXTENSIONS

Consider a two-server two-customer version of the previous model in which customers who find both servers occupied wait until a server is available. If a customer's type becomes known only when he is about to enter service and customers are admitted to service on a first come, first served basis, then the method used to prove Theorems 1 and 1′ can be utilized to yield an incredibly tedious proof of the optimality of an LIF policy for this model.

5. ACKNOWLEDGMENTS

This paper is based on Chapter 6 of the author's Ph. D. dissertation at Yale University. The author is grateful for the guidance provided by the members of his thesis committee, Matthew J. Sobel and Ward Whitt. The author also acknowledges the financial assistance of the United States Public Health Service and National Science Foundation through grants HS–00090–4 and GK 38121.

BIBLIOGRAPHY

[1] Carter, G. M., J. M. Chaiken and E. Ignall, "Response Area for Two Emergency Units," *Operations Research, 2*, 571–594 (1972).

[2] Denardo, E. V., *Dynamic Programming: Theory and Application* (Prentice-Hall, 1976).

[3] Derman, C., *Finite State Markovian Decision Processes* (Academic Press, 1970).

[4] Howard, R. A., *Dynamic Programming and Markov Processes* (M.I.T. Press, 1960).

[5] Johansen, S., "Existence of an Assignment Policy Maximizing (in the Sense of Stochastic Order) the Output from a Heterogeneous G/M/s/s Queuing System," submitted to *Operations Research.*

[6] Kotiah, T. and N. Slater, "On Two Server Poisson Queues with Two Types of Customers," *Operations Research, 21*, 597–603 (1973).

[7] Lippman, S., "Applying a New Technique in the Optimization of Exponential Queuing Systems," *Operations Research, 23*, 687–710.

[8] Rolfe, A., "The Control of a Multiple Facility, Multiple Channel Queuing System with Parallel Input Streams," Technical Report Number 22, Graduate School of Business, Stanford University (1965).

[9] Winston, W., "Optimal Operation of Congestion Systems with Heterogeneous Arrivals and Servers," Ph.D. dissertation, School of Organization and Management, Yale University (1975).

COMPUTING BOUNDS FOR THE OPTIMAL VALUE IN LINEAR PROGRAMMING

Markku Kallio*

Helsinki School of Economics
Helsinki, Finland

ABSTRACT

Consider a standard linear programming problem and suppose that there are bounds available for the decision variables such that those bounds are not violated at an optimal solution of the problem (but they may be violated at some other feasible solutions of the problem). Thus, these bounds may not appear explicitly in the problem, but rather they may have been derived from some prior knowledge about an optimal solution or from the explicit constraints of the problem.

In this paper, the bounds on variables are used to compute bounds on the optimal value when the problem is being solved by the simplex method. The latter bounds may then be used as a termination criteria for the simplex iterations for the purpose of finding a "sufficiently good" near optimal solution. The bounds proposed are such that the computational effort in evaluating them is insignificant compared to that involved in the simplex iterations. A numerical example is given to demonstrate their performance.

1. INTRODUCTION

Our purpose is to establish bounds on the optimal value of a linear programming problem. These bounds can be used as a rule for stopping when simplex iterations are carried out in order to find a near optimal solution. Our motivation in doing this is that a practical problem usually has a large number of (feasible) extreme point solutions with the objective function value relatively close to the optimal one. Therefore, also a large number of iterations are expected without significant improvement in the objective function value. We intend to avoid the computational work involved in these iterations.

The bounds can be utilized when a near optimal solution is satisfactory compared to the optimal solution. This is usually the case for linear programming models in practice. We shall also point out special applications for the branch and bound method (e.g. [6]) and for Dantzig-Wolfe decomposition [4].

Of course, any feasible solution to a (maximization) problem determines a lower bound on the optimal value. To find upper bounds we utilize the duality theory (see e.g. [1]) or, equivalently, Lagrangean relaxation [5]. A restricted dual problem is solved when evaluating an upper bound.

*This work was carried out while the author was at the European Institute for Advanced Studies in Management, Brussels, Belgium.

Essentially this requires minimization of a piecewise linear convex function. The restriction is chosen so that some computation which was already carried out by the simplex method can be utilized again. Because of these two characteristics the evaluation of an upper bound is computationally inexpensive.

2. THE PROBLEM AND ITS RESTRICTION

We consider the usual linear programming problem (LP):

find $x \in R^n$ to

$$\begin{aligned} &\text{(LP.1)} \quad \max cx \\ &\text{(LP.2)} \quad \text{s.t. } Ax=b \\ &\text{(LP.3)} \quad 0 \leqq x \leqq t, \end{aligned}$$

where $A=(a_j) \in R^{m \times n}$, $a_j \in R^m$ for all j, $b \in R^m$, $c=(c_j) \in R^n$, and $t \in R^n$. Here some of the components of the upper bound vector may be infinite. For $l=(l_j) \geqq 0$ and $u=(u_j)$ in R^n we define the problem (LP (l, u)) which differs from (LP) in that (LP.3) is replaced by:

$$l \leqq x \leqq u. \tag{1}$$

If (.) is a linear programming problem, we denote its optimal value by v (.). In this paper we shall assume that l and u are such that;

$$v(\text{LP})=v(\text{LP}(l, u)). \tag{2}$$

Our purpose in defining (LP(l, u)) is to find upper bounds on v(LP) by first finding dual suboptimal solutions for (LP(l, u)) and then applying the weak duality result (e.g. [1]) together with (2). For this purpose we think of the constraints (1) to be such that $0 \leqq l \leqq u \leqq t$ and $u < \infty$. (Note that this is not necessary for the following to be valid, but otherwise we may obtain bounds which are infinite and, therefore, useless). Thus, we think of (LP(l, u)) as being a restriction of (LP) but having the same optimal value as (LP) does. Such vectors l and u may be derived from (LP.2) and (LP.3) (for an example see Section 4), from (LP.3) alone, or even from empirical knowledge about the problem.

3. SIMPLE BOUNDS

Suppose (LP) is being solved by the simplex method (possibly combined with the upper bounding technique [2]). Denote by z the current value of the objective function (corresponding to a feasible solution for (LP)), $\bar{\lambda} \in R^m$ the current price vector for constraints (LP.2) and $\bar{c}_j \equiv c_j - \bar{\lambda} a_j$ for all j.

Thus, $\bar{c}_j$ is the reduced cost of the variable x_j if it is currently not at its upper bound. Otherwise $\bar{c}_j$ is the price associated with the upper bound of x_j. What we shall call *simple bounds* on v (LP) are given as follows:

THEOREM 1: If l and u satisfy (2) and we denote $\bar{\delta}=(\bar{\delta}_j)$ and $\bar{\mu}=(\bar{\mu}_j)$, where $\bar{\delta}_j \equiv \max\{0, \bar{c}_j\}$ and $-\bar{\mu}_j \equiv \min\{0, \bar{c}_j\}$, for all j, then

$$z \leq v(\text{LP})=v(\text{LP}(l, u)) \leq \bar{\lambda} b + \bar{\delta} u - \bar{\mu} l. \tag{3}$$

PROOF: The left hand inequality follows from the feasibility of the current solution for (LP) and the equality is valid by assumption. To prove the right hand inequality, we first state the dual of (LP(1, u)), denoting it by (D):

$$\text{find } \lambda \epsilon R^m \text{ and } \delta=(\delta_j),\ \mu=(\mu_j)\epsilon R^n \text{ to}$$

$$\min \lambda b+\delta u-\mu l \tag{D.1}$$

$$s.t.\ \lambda A+\delta-\mu \geq c \tag{D.2}$$

$$\delta,\ \mu \geq 0. \tag{D.3}$$

We now verify that $(\bar{\lambda},\ \bar{\delta},\ \bar{\mu})$ is a feasible solution for (D). Thus, the right hand inequality follows from the weak duality theorem (e.g. [1]).||

4. IMPROVED BOUNDS

We shall show how the simple bounds can be easily improved. The method used to obtain the improved bounds tends to avoid extra computation by utilizing some computation which has already been carried out by the simplex method. It will become clear that this computation would otherwise represent a major effort in computing the upper bound.

We shall first consider a class of restrictions of (D) that are easy to solve (and thereby obtain a sub-optimal solution for (D)). Let d and g be vectors in R^m and let (RD) be the problem which results from (D) when λ is restricted to the following set:

$$\{\lambda | \lambda = g+\theta d,\ \theta \epsilon R\}. \tag{4}$$

For a moment we consider g and d as being arbitrarily chosen. Later on we shall specify them in a way which makes the evaluation of the optimal value of (RD) (the improved bound) computationally simple.

We can now state (RD) as follows:

$$\text{find } \theta \epsilon R, \text{ and } \delta,\ \mu \epsilon R^n \text{ to}$$

$$\min K+p\theta+\delta u-\mu l \tag{RD.1}$$

$$\text{s.t. } h\theta+\delta-\mu \geq c^* \tag{RD.2}$$

$$\delta,\ \mu \geq 0, \tag{RD.3}$$

where

$$K \equiv gb,\ p \equiv db, \tag{5}$$

and

$$h=(h_j) \equiv dA,\ c^*=(c_j^*) \equiv c-gA. \tag{6}$$

Let $v(RD(\theta))$ be the optimal value of (RD) given θ. We shall find $v(RD)$ via minimization of $v(RD(\theta))$. The result below shows that $v(RD(\theta))$ is a concave and piecewise linear function whose value and marginal value (for any fixed θ) are readily available. Thus $v(RD(\theta))$ can be minimized by

marginal analysis: find θ such that for $\theta=\bar{\theta}$ the marginal value of $v(RD(\theta))$ vanishes or changes its sign.

THEOREM 2: $v(RD(\theta))$ is a concave and piecewise linear function of θ. If $v_+(RD(\theta))$ is the right-hand derivative of $v(RD(\theta))$ (with respect to θ), then the possible discontinuity points of $v_+(RD(\theta))$ are $\theta_1, \ldots, \theta_n$ where:

$$\theta_j=\begin{cases}-c_j^*/h_j, & \text{if } h_j\neq 0\\ \infty & \text{if } h_j=0, \text{ for all } j.\end{cases} \tag{7}$$

Furthermore,

$$v_+(RD(\theta))=p-\sum_{j\epsilon I(\theta)} u_j h_j+\sum_{j\epsilon\sim I(\theta)} l_j h_j, \tag{8}$$

and

$$v(RD(\theta))=K+p\theta+\sum_{j\epsilon I(\theta)} u_j(c_j^*-h_j\theta)-\sum_{j\epsilon\sim I(\theta)} l_j(c_j^*-h_j\theta), \tag{9}$$

where

$$I(\theta)=\{j|\theta_j\leqq\theta \text{ and } h_j<0 \text{ or } \theta_j>\theta \text{ and } h_j\geqq 0\}. \tag{10}$$

PROOF: For a fixed value of θ, (RD) decomposes into a small problem $(j(\theta))$ for each j:

find $\delta_j, \mu_j\epsilon R$ to

min $u_j\delta_j-l_j\mu_j$

s.t. $\delta_j-\mu_j\geqq c_j^*-h_j\theta$

$\delta_j, \mu_j\geqq 0$.

Let θ_j be given by (7). Then, because $0\leqq l_j\leqq u_j$,

$$v(j(\theta))=\begin{cases}u_j(c_j^*-h_j\theta) \text{ if } j\epsilon I(\theta)\\ l_j(c_j^*-h_j\theta) \text{ otherwise.}\end{cases} \tag{11}$$

Thus (because $l_j, u_j\geqq 0$) $v(j(\theta))$ is a concave and piecewise linear function (the only possible discontinuity point of its derivative being θ_j). In this notation, we have:

$$v(RD(\theta))=K+p\theta+\sum_j v(j(\theta)). \tag{12}$$

Thus, $v(RD(\theta))$ is a concave and piecewise linear function because it is a sum of a finite number of such functions. Clearly, the possible discontinuity points of its (right hand) derivative are $\theta_1, \ldots, \theta_n$. (9) follows combining (10)—(12). Then (8) follows from (9) and (10).||

When solving (RD) we first choose $\theta_i\epsilon\{\theta_j|j=1, \ldots, n\}$. If $v_+(RD(\theta))$ equals zero or changes its sign at $\theta=\theta_i$, then θ_i is optimal. Otherwise, if $v_+(RD(\theta_i))<0(>0)$, we increase (decrease) θ from θ_i to the next largest (smallest) element in $\{\theta_j|j=1, \ldots, n\}$. We continue similarly until the optimum is found. (Of course, alternative search techniques (for example Fibonacci search (e.g. [8])) can be used to find the optimal θ). Then, applying (9), we evaluate $v(RD)$. We call this the *improved upper bound.*

The following result together with Theorem 2 shows that, if each component of u is finite, then we obtain finite upper bounds on v(LP).

THEOREM 3. If (LP) is feasible, (2) holds, and $u<\infty$, then

$$-\infty<v(RD)=\inf_{\theta} v(RD(\theta))<\infty. \tag{13}$$

PROOF: By assumption, (LP(1, u)) is feasible. Therefore $v(D)>-\infty$. This implies

$$v(RD)>-\infty.$$

We verify directly that (RD) is feasible if $u<\infty$ (that is, (RD) has a feasible solution with the objective function value $<\infty$). Thus, $v(RD)<\infty$. The equality follows easily from a contradictory assumption.||

We consider now the computational effort needed in evaluating the improved upper bound. First, the constants are computed according to (5)—(7). In general, the major effort here is caused by the large number of inner products in computing the vectors h and c^*. As is the case for reduced costs (see [7]), this effort may not be insignificant when compared with the computational effort in one simplex iteration. When performing the optimality test for different values of θ, (8) is applied repeatedly. However, when one marginal value is known, another can be computed recursively from it. This yields computational savings when the simple search (described above) is applied. Finally, (9) is applied once to evaluate the bound.

We shall suggest below two rules for choosing the vectors g and d so that most of the extra computation needed to compute h and c^* is avoided. Obviously other rules can be constructed with the same advantage.

RULE A: If $\bar{\lambda}$ is the current price vector, we define $d=\bar{\lambda}$ and $g=0$. Then $h=c-(\bar{c}_j)$. and $c^*=(c_j)$. Note that now $v(RD(\theta))$ for $\theta=1$ is the simple upper bound of Theorem 1. Thus the improved bound is at least as good as the simple one in this case.

RULE B: If (d', g') is the vector pair chosen at the preceding iteration (as (d, g)) and $\bar{\theta}$ is the minimizer of $v(RD(\theta))$ in that iteration, we define $d=\bar{\lambda}$ and $g=g'+\bar{\theta}d'$. Now $v(RD(\theta))$, for $\theta=0$, equals the bound evaluated at the preceding iteration. Thus the sequence of bounds is monotonically decreasing in this case.

5. NUMERICAL EXAMPLES

We shall next give an example showing the computations needed for the simple bound and the improved bound when Rule A is applied. Thereafter we investigate by an other example the performance of the different bounds. Consider the following problem (LP):

find $x\epsilon R^6$ to

$$\max\ (5, 5, 6, 0, 0, 0)\,x$$

$$\text{s.t.} \begin{bmatrix} 1 & 2 & 1 & 1 & 0 & 0 \\ 2 & 2 & 3 & 0 & 1 & 0 \\ 4 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} x = \begin{bmatrix} 4 \\ 6 \\ 4 \end{bmatrix}$$

$$x \geq 0$$

The nonnegativity constraint together with the first equality constraint requires $x_1 \leqq 4$, $x_2 \leqq 2$, $x_3 \leqq 4$ and $x_4 \leqq 4$, and together with the second and third equality, respectively, $x_5 \leqq 6$ and $x_6 \leqq 4$. Thus, we have $l=0$ and $u=(4, 2, 4, 4, 6, 4)$ defining (LP(l, u)).

Suppose for (LP) a simplex iteration starts with the basic feasible solution $x=(0, 0, 2, 2, 0, 2)$. The corresponding price vector is $\bar{\lambda}=(0, 2, 0)$, the vector of reduced costs $(\bar{c}_j)=(1, 1, 0, 0, -2, 0)$, and the current value of the objective function (and lower bound on the optimal value) $z=12$. In order to compute the upper bound on the optimal value we choose (in the notation of Section 3) $(d, g)=(\bar{\lambda}, 0)$ according to Rule A. Then the restricted dual problem (RD) is as follows:

find $\theta \epsilon R$ and δ, $\mu \epsilon R^6$ to

min $12\,\theta+\delta(4, 2, 4, 4, 6, 4)^T$

s.t. $(4, 4, 6, 0, 2, 0)\,\theta+\delta-\mu \geqq (5, 5, 6, 0, 0, 0)$

δ, $\mu \geqq 0$.

In order to find the upper bound $v(RD)$ we first compute according to (7) $(\theta_i)=(1.25, 1.25, 1.0, \infty, 0, \infty)$. Initially we compute, according to (8), the marginal value $v_+(RD(\theta))$, say, at $\theta=\theta_3=1$: $v_+(RD(1))=-12$. Thus the optimal value of θ is not less than θ_3. Starting from θ_3, the next largest element of $\{\theta_i\}$ is $\theta_1(=\theta_2)=1.25$. We evaluate $v_+(RD(\theta_1))=12$. Thus the marginal value changes its sign at $\theta=\theta_1$. Therefore, θ_1 is optimal for (RD). By (9) we compute $v(RD)=15$. We have now $12 \leqq v(\text{LP}) \leqq 15$. The simple bound given by Theorem 1 is 18 $(=v(RD(1)))$. Thus considerable improvement is obtained by the marginal analysis approach. One can verify that $v(\text{LP})=13.8$.

As another example, a small production allocation problem with 19 constraints (in (LP.2)) and 43 variables was investigated. Based on the constraints of this problem, the bounds on the decision variables were easily derived. The optimal solution was found in 17 simplex iterations (on Phase II) and at each iteration the simple bound as well as the improved bounds according to Rules A and B were evaluated.

Figure 1 shows the bounds computed at each iteration. Notice that the current upper bound for one particular iteration is the smallest upper bound so far evaluated. Thus, the improved bounds when Rules A and B are applied perform approximately equally well and far better than the simple bounds.

Consider iteration 13, where the lower bound (the current solution value of (LP)) is 0.29 below the optimal value. The simple bound, and the improved bounds applying Rules A and B are 3.58, 0.82 and 0.72, respectively, above the optimal value. Thus Rule A implies that further iterations can improve the objective function value by no more than 1.11 (the corresponding number for Rule B being 1.01). If this satisfies ones termination criterion, the remaining four iterations, that is, 23% of Phase II iterations, can be neglected.

6. POSSIBLE APPLICATIONS AND EXTENSIONS

Linear programs in practice usually have a bounded optimal solution and often one can easily find finite bounds for each variable so that these bounds are not violated at an optimal solution of the problem. Therefore, the bounds developed here are expected to be useful wherever a near

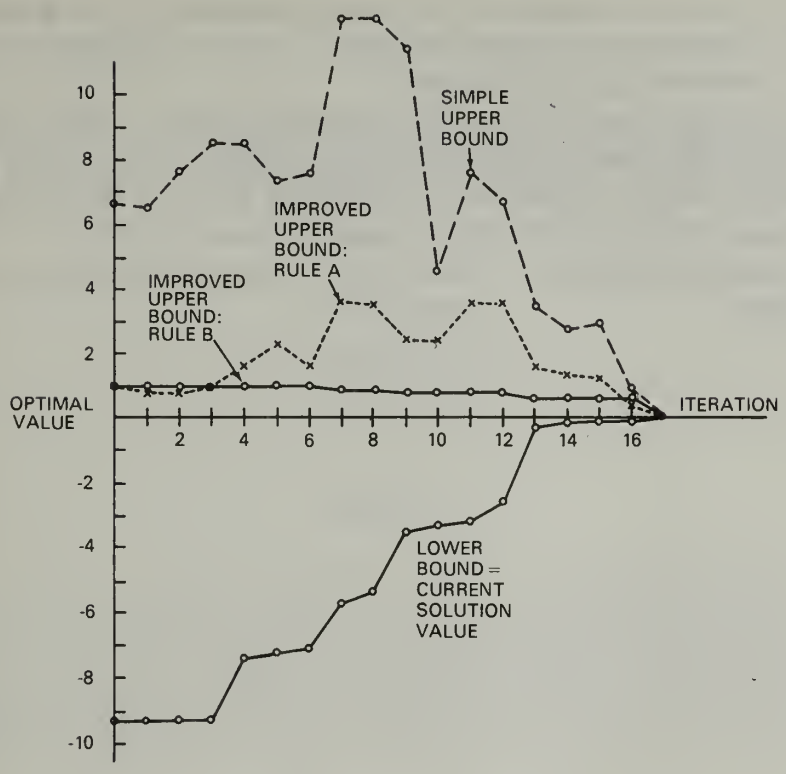


FIGURE 1. An Example of Bounds

optimal solution is satisfactory for practical purposes. Two natural applications also arise. In the branch and bound method (e.g. [6]), when a candidate problem is being maximized and for its optimal value an upper bound has been found that is less than (or equal to) the incumbent value, then the candidate problem can be fathomed. The same simple idea is applicable in Dantzig-Wolfe decomposition (e.g. [4]) when predicting whether or not a subproblem is able to create a proposal that would improve the objective function value in the master problem.

In an analogous way bounds on the optimal value can be developed when generalized upper bounds (e.g. [3]) are known. This further suggests a special application for the transportation problem.

REFERENCES

[1] Dantzig, G., *Linear Programming and Extensions* (Princeton University Press, Princeton, New Jersey, 1963).

[2] Dantzig, G., "Upper Bounds, Secondary Constraints and Block Triangularity in Linear Programming," Econometrica, *23*, 174–183 (1955).

[3] Dantzig, G. and R. Van Slyke, "Generalized Upper Bounding Techniques for Linear Programming," Journal of Computer and System Sciences, *1*, 213–226 (1967).

[4] Dantzig, G. and P. Wolfe, "Decomposition Principle for Linear Programs," Operations Research, *8*, 101–111 (1960).

[5] Geoffrion, A., "Lagrangean Relaxation for Integer Programming," in *Mathematical Programming Study 2: Approaches to Integer Programming*, ed. M. L. Balinski (North-Holland Publishing Co., Amsterdam, pp. 82–114 (1974).

[6] Geoffrion, A. and R. Marsten, "Integer Programming Algorithms: A Framework and State-of-the-Art Survey," Management Science, *12*, 456–491 (1972).

[7] Kallio, M. and E. Porteus, "Estimating Computational Effort for Linear Programming Algorithms," Research Paper No. 239, GBS, Stanford University (1975).

[8] Wagner, H., *Principles of Operations Research with Applications to Managerial Decisions* (Prentice-Hall, 1969).

SOLVING MULTICOMMODITY TRANSPORTATION PROBLEMS USING A PRIMAL PARTITIONING SIMPLEX TECHNIQUE

Jeff L. Kennington

Southern Methodist University
Dallas, Texas

ABSTRACT

This paper presents the details for applying and specializing the work of Saigal [28] and Hartman and Lasdon [16] to develop a primal partitioning code for the multicommodity transportation problem. The emphasis of the paper is in presenting efficient data structure techniques for exploiting the underlying network structure of this class of problems. Computational experience with test problems whose corresponding linear programming formulation has over 400 rows and 2,000 columns is presented.

I. INTRODUCTION

The Multicommodity Transportation Problem may be simply stated in terms of a distribution problem in which there are M suppliers (warehouses or factories), N customers (destinations), and K commodities. Each supplier, $i=1, \ldots, M$, has S_{ik} units of commodity k and each customer, $j=1, \ldots, N$, demands D_{jk} units of commodity k. Each supplier can ship units of commodity k to each destination at a shipping cost per unit of c_{ijk} (unit cost for shipping commodity k from supplier i to customer j). Further, each arc (i, j) has capacity b_{ij}. The objective is to determine which routes to use and the shipment size so that the total transportation cost of meeting demand, given supply and arc capacity constraints, is minimized. Mathematically this problem may be stated in terms of a [illegible]v structured linear program as follows:

$$\min \sum_{ijk} c_{ijk} x_{ijk} + C \sum_{jk} a_{jk} \qquad (MCTP)$$

subject to

$$\sum_j x_{ijk} + r_{ik} = S_{ik}, \text{ all } i \text{ and } k \qquad (\pi_{ik})$$

$$-\sum_i x_{ijk} - a_{jk} = -D_{jk}, \text{ all } j \text{ and } k \qquad (\pi_{M+j,k})$$

$$\sum_k x_{ijk} + s_{ij} = b_{ij}, \text{ all } i \text{ and } j$$

$$x_{ijk}, r_{ik}, a_{jk}, s_{ij} \geq 0, \text{ all } i, j, \text{ and } k, \qquad (\lambda_{ij})$$

where x_{ijk} denotes the flow of commodity k from source i to destination j, r_{ik} denotes the slack variable associated with the supply constraint for commodity k at source i, a_{jk} denotes the artificial

variable associated with the demand constraint for commodity k at destinatoin j, and s_{ij} is a slack variable associated with the capacity constraint for the arc (i, j). The Greek letters π_{ik}, $\pi_{M+j,k}$, and λ_{ij}, denote the dual variables associated with the supply, demand, and capacity constraints respectively. The cost C of artificial variables is taken to be $\Sigma_{ijk} c_{ijk}$.

1.1 Applications

Multicommodity network flow problems have been extensively studied because of their numerous applications and because of their intriguing network structure. Geoffrion and Graves [9] solved a large multicommodity warehouse location problem for Hunt-Wesson Foods, Inc. Multicommodity models have been proposed for planning studies involving urban traffic systems (see Jorgensen [22] and LeBlanc [27]) and communication systems (see White [34], and Gomory and Hu [12]). Models for solving scheduling and routing problems have been proposed by Bellmore, Bennington and Lubore [2], by White and Wrathall [33], and by Swoveland [31]. A particularly interesting application was suggested by Clarke and Surkis [5] for assigning students to schools to achieve a desired ethnic composition.

1.2 Survey of Literature

There are two basic approaches which have been employed to develop specialized techniques for multicommodity network flow problems; *decomposition* and *partitioning*. Decomposition approaches may be further characterized as *price-directive* or *resource directive*. The papers [1, 4, 6, 8, 32, 35] are all variations of price-directive decomposition while the papers [17, 29, 31] are resource-directive decomposition techniques. Partitioning procedures may be found in [13, 14, 16, 19, 28]. The only special results for the multicommodity transportation problem appeared in a recent paper by Evans, Jarvis, and Duke [7]. They showed that a necessary and sufficient condition for the constraint matrix of MCTP to be totally unimodular is that it have no more than two sources or two destinations.

1.3 Direction and Motivation of Investigation

In recent years there have been several extremely successful specializations of the primal simplex method for solving one-commodity network flow problems (see [11, 25, 30]). Primal simplex codes have been developed which are superior to the best out-of-kilter codes by a factor of at least nine to one. The primal codes are also superior to dual codes based on the same specializations.

We believe that the success of these primal codes is attributable to three key factors. First, the primal simplex technique has a computational advantage over a dual simplex technique when the problems are extremely rectangular (i.e. problems with many more columns than rows). The computational burden for a dual pivot is on the order of the number of columns whereas the burden of a primal pivot is on the order of the number of rows. Since most network flow problems are rectangular, the primal method has proven superior in computational investigations. Secondly, the simplex specializations used by these codes *strongly exploit* the underlying network structure. Finally, these implementations use efficient data structure techniques originally developed by computer scientists.

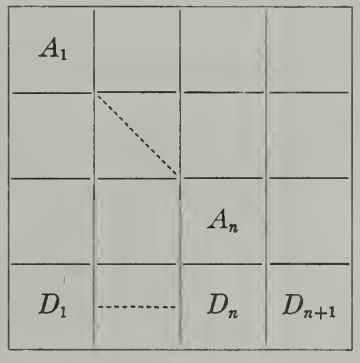
This success with one-commodity problems leads one to speculate that good results could also be obtained by extending these ideas for multicommodity problems. Unfortunately, LP bases for multicommodity problems do not exhibit the triangular property present in one-commodity problems. However, due to the underlying network structure, part of every LP basis for multicommodity problems does exhibit the triangular property and can be exploited in executing the simplex operations. That is, a sizeable part of the current basis may be stored using Johnson's [21] triple labels (or some other appropriate representation) while the remainder of the basis is stored in the usual matrix form. Hartman and Lasdon [16] have shown that the simplex operations can be performed if one carries a triangular basis of size n for each commodity, where n denotes the number of nodes in the network, and a working basis inverse whose size need never exceed the number of saturated arcs (i.e., arcs whose total flow equals arc capacity). For the multicommodity transportation problem, the triangular part of the basis has dimension $(M+N)K$. The working basis inverse varies in size with a maximum size of MN. The remaining basic columns have a single nonzero entry and are also exploited in the simplex procedure. The purpose of this study is to investigate this general approach when applied to the multicommodity transportation problem.

1.4 Notation

The notation and conventions used in this exposition are now presented. Matrices and sets are denoted by upper case Latin letters. Lower case Latin letters underlined denote column vectors. The zero vector is denoted by $\underline{0}$, and an identity matrix is denoted by I. A' and $\underline{b}'$ denote the transposition of the matrix A and the column vector $\underline{b}$ respectively. The symbol "$\equiv$" is used in place of the expression "is equivalent to" while "$\not\equiv$" is used in place of "is not equivalent to." Φ is used to denote the empty set.

II. SIMPLEX SPECIALIZATION FOR MCTP

A linear program is said to have a *block diagonal structure* if by row and column interchanges, its constraint matrix can be placed in the following form,



If the diagonal blocks, $A_1, \ldots, A_n$, each consist of a single row, then these constraints are called GUB (generalized upper bound) constraints. The general idea of partitioning block diagonal structured linear programs whose blocks have more than one row was proposed independently by Bennett [3] and Kaul [24]. These procedures were refined and specialized for multicommodity

network flow problems by Saigal [28]. This specialization involves carrying a working basis inverse whose size need never exceed the number of saturated arcs. Hartman and Lasdon [16] developed efficient procedures for updating this working basis inverse. However, neither Saigal nor Hartman and Lasdon discuss how this procedure may be efficiently implemented. Implementation is the topic of interest in this paper, and the work of Saigal and Hartman and Lasdon provides the starting point for our work.

This section presents computational devices for storing and manipulating the problem data to enhance the efficiency of a computer implementation of the primal partitioning approach. We have drawn freely from the work of Ellis Johnson [20, 21], Glover, Karney, and Klingman [10], and our own experience in solving transportation problems [25]. Our aim is to help bridge the gap between the algorithm and an efficient computational implementation.

2.1 Basis Structure

It is well known and easily proved (see [16, 26]) that by row and column interchanges, every LP basis for MCTP may be partitioned as follows:

$B^*=$

Key Columns			Nonkey Colums			
B_1			R_1			
	$\ddots$			$\ddots$		
		B_K			R_K	
P_1	⋯	P_K	T_1	⋯	T_K	
S_1	⋯	S_K	U_1	⋯	U_K	I

where $\det(B_k)\neq 0$ for $k=1, \ldots, K$. That is, B_l is a basis for the following transportation problem:

$$\min \sum_{ij} c_{ijl}x_{ijl}+C\sum_{j} a_{jl} \qquad (TP_l)$$

$$\text{st.} \sum_{j} x_{ijl}+r_{il}=S_{il}, \text{ all } i$$

$$-\sum_{i} x_{ijl}-a_{jl}=-D_{jl}, \text{ all } j$$

$$x_{ijl}\geq 0.$$

The definitions which follow are used to characterize a basis of a transportation problem in terms of a graph. The motivation is to use a set of graphs (one for each commodity) to efficiently carry out the simplex operations involving $B_1, \ldots, B_K$. The notation is identical to that used by Johnson [20].

A *graph* G is a finite set V of vertices (nodes) $v_1, \ldots, v_k$, and a set E of unordered pairs of vertices, $e_p=(v_i, v_j)$, called edges, (arcs). A *path* in a graph is a sequence of vertices and distinct edges $(v_1, e_1, v_2, e_2, \ldots, v_{k-1}, e_{k-1}, v_k)$, such that $e_i=(v_i, v_{i+1})$. A *simple path* is a path with distinct vertices, and a *cycle* is a simple path together with an edge from the beginning to the end of the path. A *connected graph* has at least one path between every pair of vertices. A connected graph with no cycles is called a *tree* and a graph consisting of one or more unconnected trees is called a *forest*. A *spanning subgraph* of G is a graph with the same vertex set as G, and a *spanning* forest of G is a forest which is a spanning subgraph.

Let A_0 denote the constraint matrix of TP_l. Then A_0 can be partitioned as $A_0=[A, U]$, where A is a node-arc incidence matrix and U is a diagonal matrix with ones and minus ones along the diagonal. The rows of A_0 can be viewed as representing vertices while the columns represent edges of the graphical representation of TP_l. Let F denote the graph associated with TP_l. Columns of U represent edges incident to a single vertex and are called slack or artificial arcs corresponding to the slack or artificial variables. A tree with one slack or artificial arc incident to some vertex of the tree is called a *rooted tree*, and the slack or artificial arc is called the *root of the tree*. A forest of the graph corresponding to TP_l, with each tree having one root is called a *rooted forest*. Let B_l be a basis of A_0. Then the *subgraph* generated by the basis B_l called F_{B_l} consists of all vertices of F, edges corresponding to columns of A in B_l, and roots from columns of U in B_l. Johnson [20] has shown that B_l is a basis of A_0 if and only if F_{B_l} is a rooted spanning forest of F.

Since bases for TP_l have a graphical characterization, this graph can be stored rather than the matrix. A particularly efficient storage scheme for matrices of this type involves three labels for each node (see [21]). We propose to use this scheme for storing each of the bases $B_1, \ldots, B_K$. This allows the revised simplex operations involving $B_1, \ldots, B_K$ to be performed via tracing labels through a graph, rather than by matrix multiplication.

It will now be shown that by column operations, B^* can be converted into block diagonal form. Let the matrix L be defined as follows:

$$L=\begin{array}{|c|c|c|c|}\hline & -B_1^{-1}R_1 & & \\ \cline{2-4} I & & & \\ \cline{3-4} & & & -B_K^{-1}R_K \\ \hline & I & & \\ \hline & & & I \\ \hline \end{array}$$

Post multiplying B^* by L yields

$$B^*L=\begin{array}{|c|c|c|c|c|}\hline B_1 & & & & \\ \hline & \diagdown & & & \\ \hline & & B_K & & \\ \hline P_1 & \cdots & P_K & Q & \\ \hline S_1 & \cdots & S_K & D & I \\ \hline\end{array}$$

where

$$\begin{array}{c} q\left\{\right. \\ w\left\{\right. \end{array}\begin{array}{|c|c|}\hline Q & \\ \hline D & I \\ \hline\end{array}=\begin{array}{|c|c|c|}\hline & T_k-P_kB_k^{-1}R_k & \\ \hline & U_k-S_kB_k^{-1}R_k & I \\ \hline\end{array}$$

It will now be shown that Q and D can be easily generated if the bases, $B_1, \ldots, B_K$, and the nonkey columns are known. Let x_{ijk} be a nonbasic or nonkey column incident on nodes n_l and n_r in F_{B_k} where n_l is a source and n_r is a destination. Let P'_L denote the unique simple path in F_{B_k} from n_l to the root of the tree associated with n_l, including the root. All arcs in the path which are traversed in reverse direction (i.e., from destination to source) are said to have reverse orientation. Let P'_R be the unique simple path from the root of the tree associated with n_r to n_r including the root. Let

$$P_L=P'_L-(P'_L\cap P'_R) \text{ and } P_R=P'_R-(P'_L\cap P'_R).$$

For the slack variable, r_{ik}, the definition of P'_L is as described above and $P'_R=\phi$. Let a^* denote the column of TP_k corresponding to x_{ijk} or r_{ik}. The following proposition indicates a direct means for constructing $B_k^{-1}a^*$.

PROPOSITION 1: The i^{th} component of $\underline{y}^*=B_k^{-1}\underline{a}^*$ is determined as follows: $+1$, if the arc corresponding to the i^{th} column of B_k is in $P_L\cup P_R$ with normal orientation (i.e., from source to destination), -1, if the arc corresponding to the i^{th} column of B_k is in $P_L\cup P_R$ with reverse orientation, 0, otherwise. A proof of Proposition 1 may be found in [25].

Let $T=\{t_1, \ldots, t_q\}$ where $t_l=(r, s)$ implies that the l^{th} row of Q corresponds to the constraint $\Sigma_k x_{rsk}+s_{rs}=b_{rs}$. That is, T is the index set for the rows in Q. Let x_r associated with commodity k be some nonkey variable corresponding to the p^{th} column of Q. Let P_L and P_R be defined as above for the nonkey variable x_r. Then the p^{th} column of Q is easily constructed using the below proposition.

PROPOSITION 2: The i^{th} component of the p^{th} column of Q is as follows:

1, if $x_r \equiv x_{uvk}$ and $t_i=(u, v)$,

1, if $x_r \not\equiv x_{uvk}$, $t_i=(u, v)$, and the arc corresponding to x_{uvk} is an element of $P_L \cup P_R$ with reverse orientation,

-1, if $x_r \not\equiv x_{uvk}$, $t_i=(u, v)$, and the arc corresponding to x_{uvk} is an element of $P_L \cup P_R$ with normal orientation,

0, otherwise.

PROOF: Let $\underline{a}^*$ be the column of TP_k corresponding to x_r. Then the columns of Q associated with commodity k take the form $T_k - P_k B_k^{-1} R_k$ where $\underline{a}^*$ is a column of R_k. Suppose $\underline{a}^*$ is the j^{th} column of $T_k - P_k B_k^{-1} R_k$ and is denoted by

$$T_k^j - P_k B_k^{-1} \underline{a}^* = \begin{bmatrix} T_{1k}^j - P_{1k} B_k^{-1} \underline{a}^* \\ T_{2k}^j - P_{2k} B_k^{-1} \underline{a}^* \\ \vdots \\ T_{qk}^j - P_{qk} B_k^{-1} \underline{a}^* \end{bmatrix}$$

Let $i \in \{1, \ldots, q\}$ and suppose $t_i=(u,v)$. Then the i^{th} component of the p^{th} column of Q corresponds to arc $t_i=(u,v)$ and is given by $T_{ik}^j - P_{ik} B_k^{-1} \underline{a}^*$.

CASE 1. Suppose $x_r \equiv x_{uvk}$. Then $T_{ik}^j=1$ and $P_{ik}=\underline{0}'$. Hence, $T_{ik}^j - P_{ik} B_k^{-1} \underline{a}^*=1$.

CASE 2. Suppose $x_r \not\equiv x_{uvk}$ and the arc corresponding to x_{uvk} is in $P_L \cup P_R$ with reverse orientation. Suppose the arc corresponding to x_{uvk} appears in the t^{th} column of B_k. Then

$$T_{jk}^t - P_{ik} B_k^{-1} \underline{a}^* = 0 - [\underline{0}' \underset{\substack{\uparrow \\ t^{\text{th}}}}{1} \underline{0}'] \begin{bmatrix} \\ -1 \\ \\ \end{bmatrix} \leftarrow t^{\text{th}} = 1,$$

where $B_k^{-1} \underline{a}^*$ is given by Proposition 1.

CASE 3. Suppose $x_r \not\equiv x_{uvk}$ and the arc corresponding to x_{uvk} is in $P_L \cup P_R$ with normal orientation. Suppose the arc corresponding to x_{uvk} is in the t^{th} column of B_k. Then

$$T_{ik}^j - P_{ik} B_k^{-1} \underline{a}^* = 0 - [\underline{0}' \underset{\substack{\uparrow \\ t^{\text{th}}}}{1} \underline{0}'] \begin{bmatrix} \\ 1 \\ \\ \end{bmatrix} \leftarrow t^{\text{th}} = -1$$

where $B_k^{-1} \underline{a}^*$ is given by Proposition 1.

CASE 4. Suppose $x_r \not\equiv x_{uvk}$ and the arc corresponding to x_{uvk} is not in the set $P_L \cup P_R$. Suppose P_{ik} is $\underline{0}'$. Then clearly, the i^{th} component of the p^{th} column of Q is zero. Suppose P_{ik} has a 1 in the t^{th} position. Then

$$T_{ik}^j - P_{ik} B_k^{-1} \underline{a}^* = 0 - [\underline{0}' \underset{\substack{\uparrow \\ t^{\text{th}}}}{1} \underline{0}'] \begin{bmatrix} \\ 0 \\ \\ \end{bmatrix} \leftarrow t^{\text{th}} = 0 \tag{1}$$

This completes the proof of proposition 2.

Let $Z=\{z_1, z_2, \ldots, z_w\}$ where $z_l=(r, s)$ implies that the l^{th} row of D corresponds to the constraint $\sum_k x_{rsk}+s_{rs}=b_{rs}$. That is, Z is the index set for the rows of D. Using the definitions for x_r, P_L, and P_R given above, the following proposition shows how columns of D may be constructed.

PROPOSITION 3. The i^{th} component of the p^{th} column of D is as follows:

1, if $x_r \equiv x_{uvk}$ and $z_i=(u, v)$,

1, if $x_r \not\equiv x_{uvk}$, $z_i=(u, v)$, and the arc corresponding to x_{uvk} is an element of $P_L \cup P_R$ with reverse orientation,

-1, if $x_r \not\equiv x_{uvk}$, $z_i=(u, v)$, and the arc corresponding to x_{uvk} is an element of $P_L \cup P_R$ with normal orientation,

0, otherwise.

The proof of Proposition 3 is identical to that given for proposition 2 with the index set Z replacing T. It is now shown that the results of the above propositions can be used to specialize the primal simplex method.

2.2 Data Requirements

Assume that at the beginning of an iteration, the following quantities are explicitly stored.

(i) B_k, $k=1, \ldots, k$, each stored using Johnson's [21] three labels,

(ii) the $q \times q$ matrix Q^{-1},

(iii) the dual variables, π_{ik}, $\pi_{M+j,k}$, and λ_{ij}, and

(iv) the index sets T and Z.

Using the above data it will now be shown that the revised simplex operations can be efficiently performed.

2.3 Pricing

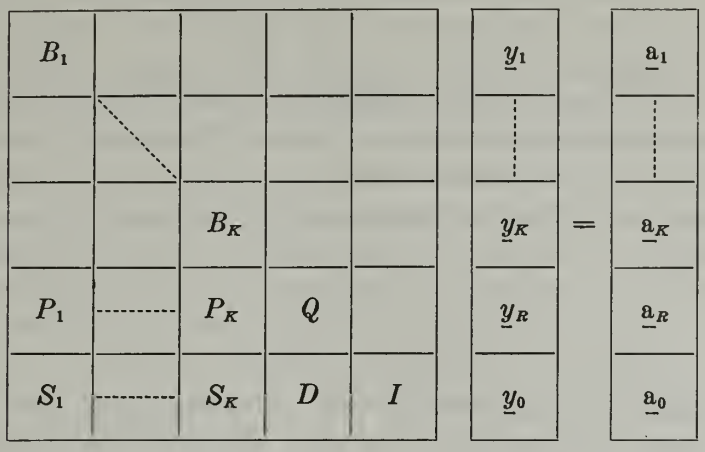
The relative cost factors are simply $c_{ik}{}^1=\pi_{ik}$, for slack variable r_{ik}, $c_{ijk}{}^2=\pi_{ik}-\pi_{M+j,k}+\lambda_{ij}-c_{ijk}$, for variable x_{ijk}, and $c_{ji}{}^3=\lambda_{ij}$, for the slack variable s_{ij}. The optimality condition is max $[c_{ik}{}^1, c_{ijk}{}^2, c_{ij}{}^3] \leq 0$ for all i, j, k. Any variable with revised cost greater than zero is a candidate to enter the basis.

2.4 Updating a Column, Selecting the Leaving Variable, and Updating Flows

Consider some column $\underline{a}$ of the constraints for MCTP. The updated column $\underline{g}$ is formed by solving the system $B^*\underline{g}=\underline{a}$. Let $\underline{y}=L^{-1}\underline{g}$ or

$$L\underline{y}=\underline{g}. \tag{2}$$

Then $B^*L\underline{y}=\underline{a}$ can be solved for $\underline{y}$ and $\underline{g}$ obtained from (2). By a suitable partitioning $B^*L\underline{y}=\underline{a}$ is



Note that at most one of the vectors $\underline{a}_1, \ldots, \underline{a}_K$ will be nonzero. If all are $\underline{0}$, then $\underline{a}$ must correspond to some slack variable, s_{ij}. If one is nonzero, denote it by $\underline{a}_l$. Then the above system yields

$$\underline{y}_k=\underline{0}, \text{ for } k=1, \ldots, K; k\neq l \tag{3}$$

$$\underline{y}_l=B_l^{-1}\underline{a}_l \tag{4}$$

$$\underline{y}_R=Q^{-1}[\underline{a}_R-P_l\underline{y}_l] \tag{5}$$

$$\underline{y}_0=\underline{a}_0-S_lB_l^{-1}\underline{a}_l-D\underline{y}_R \tag{6}$$

The equations (2)—(6) imply

$$\underline{g}_R=Q^{-1}[\underline{a}_R-P_l\underline{y}_l] \tag{7}$$

$$\underline{g}_0=\underline{a}_0-S_l\underline{y}_l-D\underline{g}_R \tag{8}$$

$$\underline{g}_k=-B_k^{-1}R_k\underline{g}_{R_k}, \text{ for } k=1, \ldots, K \text{ and } k\neq l \tag{9}$$

$$\underline{g}_l=\underline{y}_l-B_l^{-1}R_l\underline{g}_{R_l} \tag{10}$$

where $\underline{g}_R$ is partitioned into $[\underline{g}_{R_1}, \ldots, \underline{g}_{R_K}]$ and each partition has the same dimension as the corresponding R_k. We are concerned with an efficient procedure for making the computations (7)—(10). Let x_E denote the variable corresponding to $\underline{a}$ and suppose there exist an $\underline{a}_l\neq\underline{0}$. Consider the following sequence of steps.

STEP 1.: Determine $\underline{y}_l=B_l^{-1}\underline{a}_l$.

Proposition 1 is used to construct $\underline{y}_l$.

STEP 2: Calculate $\underline{g}_R = Q^{-1}[\underline{a}_R - P_l \underline{y}_l]$.

If $x_E \equiv r_{il}$ for some i, then $\underline{a}_R = \underline{0}$. Otherwise $x_E \equiv x_{uvl}$ for some u and v and $\underline{a}_R$ is either a unit vector or $\underline{0}$. If there exists a $t_p = (u, v)$ for some p, then $\underline{a}_R$ is a unit vector with a one in the p^{th} component. Otherwise, $\underline{a}_R = \underline{0}$. Attention is now turned to the calculation of the q component vector $P_l \underline{y}_l$. P_l is a $q \times M+N$ matrix whose elements are one or zero. The rows of P_l correspond to the index set T, and the columns correspond to columns of B_l. Further, each row of P_l can have at most one nonzero entry. Suppose $t_p = (u, v)$. The p^{th} row of P_l will have a nonzero entry if and only if the column corresponding to x_{uvl} appears in B_l. If x_{uvl} appears in the r^{th} column of B_l, then P_l will have a one in the $(p, r)^{\text{th}}$ element. Otherwise the p^{th} row will be $\underline{0}'$. Hence, the vector $P_l \underline{y}_l$ can be efficiently constructed as $\underline{y}_l$ is being constructed. That is, as the nonzero components of $\underline{y}_l$ are constructed, a search over the index set T can be used to construct the components of $P_l \underline{y}_l$. Once $\underline{a}_R$ and $P_l \underline{y}_l$ have been constructed, $\underline{g}_R$ is obtained by the matrix multiplication

$$Q^{-1}[\underline{a}_R - P_l \underline{y}_l].$$

STEP 3: Calculate $\underline{g}_0 = \underline{a}_0 - S_l \underline{y}_l - D\underline{g}_R$.

If $x_E \equiv r_{il}$ for some i, then $\underline{a}_0 = \underline{0}$. Otherwise $x_E \equiv x_{uvl}$ for some u and v and $\underline{a}_0$ is either a unit vector or $\underline{0}$. If there exists a $z_p = (u, v)$ for some p, then $\underline{a}_0$ is a unit vector with a one in the p^{th} component. Otherwise, $\underline{a}_0 = \underline{0}$. Let us now investigate the calculation of $S_l \underline{y}_l$, Note that S_l has the same form as P_l except the index set T is replaced by the index set Z. Hence $S_l \underline{y}_l$ can be developed during the construction of $\underline{y}_l$ by a search over the index set Z. The components of D can be constructed using Proposition 3. Hence the vector $D\underline{g}_R$ is easily obtained once $\underline{g}_R$ is determined. Finally the three vectors are combined to yield $\underline{g}_0$.

STEP 4: Determine $\underline{g}_k = -B_k^{-1} R_k \underline{g}_{Rk}$, for $k = 1, \ldots, K$ and $k \neq l$.

Proposition 1 can be used to construct columns of $-B_k^{-1} R_k$ and this matrix can be multiplied by $\underline{g}_{Rk}$. If $x_E \equiv s_{uv}$ for some (u, v), then several other simplifications arise.

STEP 5: Determine $\underline{g}_l = \underline{y}_l - B_l^{-1} R_l \underline{g}_{Rl}$

Once the updated column has been determined, the usual primal simplex ratio test is used to select the leaving variable. Given the leaving variable and the updated column, the flows are easily updated in the usual manner.

2.5 Updating Bases, $B_1, \ldots, B_K$, Working Basis Inverse, and Dual Variables

Hartman and Lasdon [16] have developed an excellent procedure for updating both Q^{-1} and λ_{ij}'s after each pivot. All updating required for the bases, $B_1, \ldots, B_K$, can be efficiently performed using the Augmented Predecessor Index Method [10]. We now address the question of updating the π_{ik}'s. Recall that for all basic columns, x_{ijk}, $\pi_{ik} - \pi_{M+j,\,k} + \lambda_{ij} - c_{ijk} = 0$, for all basic columns, r_{ik}, $\pi_{ik} = 0$, while for basic columns, a_{jk}, $-\pi_{M+j,\,k} = C$. Hence, π_{ik}'s and $\pi_{M+j,\,k}$'s for all roots are determined. Then for all arcs incident to a root node, the rule $\pi_{ik} - \pi_{M+j,\,k} - \lambda_{ik} = c_{ijk}$ can be used

to determine the π not corresponding to the root node. All of the π's are uniquely determined iteratively because of a rooted tree being connected and having no cycles.

2.6 Reinversion of Working Basic Inverse and Updating Dual Variables

Since Q^{-1} and the λ's are updated at each pivot rather than recalculated, the accumulation of roundoff error eventually forces Q^{-1} and λ to be recalculated from the original problem data. Proposition 2 can be used to develop an accurate Q and then an inversion routine can be used to calculate Q^{-1}.

To find the dual variables associated with the basis B^*, we need to solve the system

$$[\underline{\pi}', \underline{\lambda}']B^* = \underline{c}_{B^*}$$

where $\underline{c}'_{B^*}$ denotes the cost associated with the basic variables. Partition $\underline{c}'_{B^*}$ as $[\underline{c}'_{B_1}, \ldots, \underline{c}'_{B_K}, \underline{c}'_{R_1}, \ldots, \underline{c}'_{R_K}, \underline{0}']$ where $\underline{c}'_{R_k}$ has the same dimension as R_k. By multiplying on the right by L we have the system: $[\underline{\pi}' \underline{\lambda}]B^*L = \underline{c}'_{B^*}L$ or

$\underline{\pi}'_1$	----------	$\underline{\pi}'_K$	$\underline{\lambda}'_R$	$\underline{\lambda}'_0$

B_1				
	⋱			
		B_K		
P_1	----------	P_K	Q	
S_1	----------	S_K	D	I

$$= \left| \underline{c}'_{B_1} \right| \cdots\cdots \left| \underline{c}'_{B_K} \right| \left| \underline{c}'_{R_1} - \underline{c}'_{B_1} B_1^{-1} R_1 \right| \cdots\cdots \left| \underline{c}'_{R_K} - \underline{c}'_{B_K} B_K^{-1} R_K \right| \underline{0}'$$

Then $\underline{\lambda}'_0 = \underline{0}'$ and

$$\underline{\lambda}'_R = (\underline{c}'_{R_1} - \underline{c}'_{B_1} B_1^{-1} R_1, \ldots, \underline{c}'_{R_K} - \underline{c}'_{B_K} B_K^{-1} R_K) Q^{-1}.$$

Proposition 1 is used to construct columns of $B_k^{-1}R_k$ for each $k=1, \ldots, K$. Next the components $\underline{c}'_i - \underline{c}'_{B_i} B_i^{-1} R_i$ are determined and finally the matrix multiplication by Q^{-1} yields $\underline{\lambda}_R$.

III. COMPUTATIONAL EXPERIENCE

The primal simplex specialization as described in Section 2 has been coded and was used to obtain the experience reported in Tables 1, 2, 3, and 4. The code is wirtten entirely in FORTRAN and has been tested on a CDC Cyber 72 and a CDC 6600 using the FTN compiler. The object program and data are held incore with all data in floating point mode. The rooted spanning forests are stored using Johnson's [21] triple labels. The Augmented Predecessor Index Method (see [10])

is used for updating the forests associated with each commodity. The working basis inverse (Q^{-1}) is updated using the procedure of Hartman and Lasdon [16]. The code consists of approximately sixteen hundred FORTRAN statements organized into a main program and fifteen subroutines. All data files appear in common blocks so that there is a minimum of time loss due to the subroutine structure. The storage requirement is $MNK+3MN+7(M+N)K+L^2+8L+8000$ where L is the maximum size of the working basis.

Experimental tests indicated that there was no significant difference in the code's running speed when the dual variables were recalculated at the end of a pivot rather than updating those that change. The experience reported in Tables 1, 2, and 3 is for a code that recalculated the dual variables at the end of each pivot. The variables r_{ik}, a_{jk}, s_{ij} provided the starting bases for all problems.

Table 1 presents a summary of computational time as a function of the pricing rule used. The *first positive evaluator rule* implies that the first nonbasic variable encountered with revised cost greater than zero is selected as the entering variable. Pricing all nonbasic variables and entering the one with largest revised cost is the *largest evaluator rule*. The *largest evaluator in a block rule* involves holding the commodity subscript, k, constant and pricing all nonbasic variables associated with that commodity. If all revised costs are negative, k is incremented and the procedure repeated. If a nonbasic variable prices out greater than zero for some k, then the nonbasic variable associated with commodity k with largest revised cost is selected as the entering variable. Table 2 presents the percentage of total time spent in the pricing operation for each rule.

The coefficients for the test problems were generated using uniform distributions with predetermined upper and lower limits. The arc costs were uniformly distributed over the interval [0, 100] while the supplies and demands were distributed over [100, 300]. The arc capacities were distributed over the interval [200, 600].

For our in-core code the largest evaluator rule proved superior. However, for large problems which can not be handled in-core, the largest evaluator in a block rule appears quite attractive. Additional experimental work with different partial pricing strategies is needed in order to obtain the best performance from the primal partitioning procedure.

Table 3 presents our computational experience in comparing the APEX III System with the primal partitioning code. APEX III is Control Data Corporation's production linear programming system. APEX III has an in-core optimizer which uses the product form of the inverse. Only the distinct elements of the original constraints and the eta vectors are stored explicitly. Plus and minus ones are handled separately to avoid trival multiplications and divisions by these values. The APEX system was run twice with each of the test problems. In the first run, identity matrices were used for starting bases while for the second run starting bases were crashed. However, these different starting bases produced optimal solutions in approximately the same time. The primal partitioning code used identity matrices for the starting bases for all problems.

It is anticipated that a production code would store only the nonzero elements of the working basis inverse. It is also conceivable that one could follow the suggestion of Kalan [23] and only store the distinct elements. Table 4 presents, the characteristics of the working basis inverse at optimality for several problems of various sizes. Note that a large proportion of the nonzero elements are plus and minus ones and the number of distinct elements is quite small. It appears that these characteristics are due to the fact that the elements of Q are from $\{-1,0,1\}$.

TABLE 1. *Comparison of pricing rules*

Size $M=N=K$	Equivalent LP size		First positive evaluator				Largest evaluator				Largest evaluator in a block			
	Rows	Columns	Number problems	Average time (minutes)	Average number pivots	Average time per pivot (seconds)	Number problems	Average time (minutes)	Average number pivots	Average time per pivot (seconds)	Number problems	Average time (minutes)	Average number pivots	Average time per pivot (seconds)
7	147	490	5	4. 3	4381	0. 06	5	0. 3	285	0. 06	5	0. 3	331	0. 05
8	192	704	2	11. 6	9340	0. 07	5	0. 8	465	0. 10	5	0. 7	612	0. 07
9	243	972					5	1. 7	798	0. 13	5	1. 9	1150	0. 10
10	300	1300					5	3. 0	1053	0. 17	5	4. 2	1809	0. 14
11	363	1694					4	6. 0	1539	0. 24	3	8. 7	2799	0. 19
12	432	2160					2	13. 4	2288	0. 35	1	22. 0	4924	0. 27
13	507	2704					1	25. 6	3449	0. 45				

TABLE 2. *Percentage of total time devoted to pricing with different rules*

Size $M=N=K$	Equivalent LP size		First positive evaluator			Largest evaluator			Largest evaluator in a block		
	Rows	Columns	Number problems	Total time (minutes)	Percent time for pricing	Number problems	Total time (minutes)	Percent time for pricing	Number problems	Total time (minutes)	Percent time for pricing
7	147	490	5	21. 6	2	5	1. 5	29	5	1. 3	7
8	192	704	2	23. 2	2	5	3. 8	31	5	3. 6	5
9	243	972				5	8. 4	29	5	9. 7	5
10	300	1300				5	15. 1	28	5	20. 9	4
11	363	1694				4	24. 1	27	3	26. 0	4
12	432	2160				2	26. 8	23	1	22. 0	3
13	507	2704				1	25. 6	23			

TABLE 3. *Comparison of the primal partitioning code with APEX III*

Problem description		Equivalent LP size		APEX III		Primal partitioning			Time ratio [B]/[C]
Number	Size ($M=N=K$)	Rows ($3K^2$)	Columns (K^3+3K^2)	[A] CDC 6600 identity start (seconds)	[B] CDC 6600 crash start (seconds)	[C] CDC 6600 identity start (seconds)	[D] CDC Cyber 72 identity start (seconds)	Number of pivots	
1	8	192	704	18	18	13	24	404	1. 38
2	↑	↑	↑	18	18	17	32	525	1. 06
3	↓	↓	↓	18	18	17	33	528	1. 06
4				16	16	13	24	421	1. 23
5	8	192	704	26	25	23	45	579	1. 09

TABLE 4. *Characteristics of* Q^{-1}

Observation No.	Problem size $M=N=K$	Size of Q^{-1} optimality	Density (percent)	Number nonzeroes	Number minus ones	Number plus ones	Number distinct elements
1	7	10×10	59	59	13	14	6
2	7	11×11	18	22	10	12	0
3	8	12×12	16	23	12	11	0
4	8	14×14	15	30	19	11	0
5	8	14×14	25	48	15	23	2
6	9	13×13	19	32	13	19	0
7	9	16×16	28	71	37	34	0
8	9	16×16	31	80	32	34	2
9	10	21×21	25	110	19	19	4
10	10	24×24	25	139	55	64	3

IV. SUMMARY AND CONCLUSIONS

In this study we have presented computational approaches for developing an efficient implementation of the primal partitioning simplex method for solving multicommodity transportation problems. These approaches can be easily extended to more general multicommodity network flow models and we only used the transportation structure due to the simplicity of this model. This study has indicated the viability of solving large multicommodity network flow problems via the primal partitioning technique. Our computational experience indicates that large problems in this class should be solvable in-core. That is, only modest storage requirements are needed for the working basis inverse (Q^{-1}) if it is packed and only the distinct elements are stored. Finally this study points out the ease in which the working basis can be regenerated when reinversion is required.

ACKNOWLEDGMENTS

The author wishes to express his appreciation to Gilbert Robertson of Plastics Manufacturing Company and Sarah Shipley of Control Data Corporation for helping with the computer runs involving the APEX system. Appreciation is also extended to Control Data Corporation for providing the author with free computer time and access to APEX. The author is also grateful to Mustafa Abdulaal for his helpful comments.

NOTE

This experimental code may be obtained free of charge by writing directly to the author.

BIBLIOGRAPHY

[1] Bazaraa, M. S., "An Infeasibility Pricing Algorithm for the Multicommodity Minimal Cost Flow Problem," Working Paper, Industrial and Systems Engineering Department, Georgia Institute of Technology (1973).

[2] Bellmore, M., G. Bennington and S. Lubore, "A Multivehicle Tanker Scheduling Problem," Transportation Science, *5*, 36–47 (1971).

[3] Bennett, J. M., "An Approach to Some Structured Linear Programming Problems," Operations Research, *14*, 636–645 (1966).

[4] Chen, H. and C. G. DeWald, "A Generalized Chain Labelling Algorithm for Solving Large Multicommodity Flow Problems," Computers and Operations Research, *1*, 437–465 (1974).

[5] Clark, S. and J. Surkis, "An Operations Research Approach to Racial Desegregation of School Systems," Socio-Economic Planning Science, *1*, 295–272 (1968).

[6] Cremeans, J. E., R. A. Smith and G. R. Tyndall, "Optimal Multicommodity Network Flows with Resource Allocation," Naval Research Logistics Quarterly, *17*, 269–280 (1970).

[7] Evans, J. R., J. J. Jarvis and R. A. Duke, "Matroids, Unimodularity, and the Multicommodity Transportation Problem," Presented at the Joint National Meeting of ORSA/TIMS, Chicago (1975).

[8] Ford, L. R. and D. R. Fulkerson, "A Suggested Computation for Maximal Multicommodity Network Flows," Management Science, *5*(1), 97–101 (1958).

[9] Geoffrion, A. M. and G. W. Graves, "Multicommodity Distribution System Design By Benders Decomposition," Management Science, *20*(5), 822–844 (1974).

[10] Glover, F., D. Karney and D. Klingman, "The Augmented Predecessor Index Method for Locating Stepping-Stone Paths and Assigning Dual Prices in Distribution Problems," Transportation Science, *6*, 171–179 (1972).

[11] Glover, Fred, D. Karney, D. Klingman and A. Napier, "A Computational Study on Start Procedures, Basis Change Criteria, and Solution Algorithms for Transportation Problems," Management Science, *20*(5), 793–813 (1974).

[12] Gomory, R. E. and T. C. Hu, "Multi-Terminal Network Flows," Journal of the Society for Industrial and Applied Mathematics, *9*(4), 551–570 (1961).

[13] Graves, G. W. and R. D. McBride, "A Dynamic Factorization Algorithm for General Large Scale Linear Programming Problems," Presented at ORSA/TIMS National Meeting in San Juan, Puerto Rico (1974).

[14] Grigoriadis, M. D. and W. W. White, "A Partitioning Algorithm for the Multicommodity Network Flow Problem," Mathematical Programming *3*, 157–177 (1972).

[15] Hartman, J. K. and L. S. Lasdon, "A Generalized Upper Bounding Method for Doubly Couple Linear Programs," Naval Research Logistics Quarterly, *17*(4), 411–429 (1970).

[16] Hartman, J. K. and L. S. Lasdon, "A Generalized Upper Bounding Algorithm for Multicommodity Network Flow Problems," Networks, *1*, 333–354 (1972).

[17] Held, M., P. Wolfe and H. Crowder, "Validation of Subgradient Optimization," Mathematical Programming, *6*, 62–88 (1974).

[18] Jarvis, J. J., and P. D. Keith, "Multi-Commodity Flows with Upper and Lower Bounds," Presented to the Joint National Meeting of ORSA and TIMS in Boston (Spring 1974).

[19] Jewell, W. S., "A Primal-Dual Multi-Commodity Flow Algorithm," ORC Report 66–24, University of California, Berkeley (1966).

[20] Johnson, Ellis L., "Programming in Networks and Graphs," Operations Research Center Report No. 65–1, University of California, Berkely, (1965).

[21] Johnson, Ellis L., "Networks and Basic Solutions," Operations Research, *14*(4), 619–623 (1966).

[22] Jorgensen, N. O., "Some Aspects of the Urban Traffic Assignment Problem," Graduate Report, The Institute of Transportation and Traffic Engineering, University of California, Berkeley (1963).

[23] Kalan, James E., "Aspects of Large-Scale In-Core Linear Programming," Proceedings of the 1971 Annual Conference of the ACM, Chicago, Ill., 304–313.

[24] Kaul, R. N., "An Extension of Generalized Upper Bounded Techniques for Linear Programming," ORC 65–27, Operations Research Center, University of California, Berkeley (1965),

[25] Langley, R. W., J. L. Kennington and C. M. Shetty, "Efficient Computational Devices for the Capacitated Transportation Problem," Naval Research Logistics Quarterly, *21*(4), 637–647, (1974).

[26] Lasdon, Leon S., Optimization Theory for Large Systems (Macmillan Company, New York. 1970).

[27] LeBlanc, L. J., "Mathematical Programming Algorithms for Large Scale Network Equilibrium and Network Design Problems," unpublished dissertation, Dept. of Industrial Engineering and Management Sciences, Northwestern University (1973).

[28] Saigal, Romesh, "Multicommodity Flows in Directed Networks," ORC Report 66–24, University of California, Berkeley (1966).

[29] Sakarovitch, Michel, "The Multi-Commodity Maximal Flow Problem," ORC Report 66–25, University of California, Berkeley (1966).

[30] Srinivansan, V. and G. L. Thompson, "Benefit-Cost Analysis of Coding Techniques for the Primal Transportation Algorithm," Journal, Association for Computing Machinery *20*, 194–213 (1973).

[31] Swoveland, Cary, "Decomposition Algorithms for the Multi-Commodity Distribution Problem," Working Paper No. 184, Western Management Science Institute, UCLA (1971).

[32] Tomlin, J. A., "Minimum-Cost Multicommodity Network Flows," Operations Research, *14*(1), 45–51 (1966).

[33] White, W. W. and E. Wrathall, "A System For Railroad Traffic Scheduling," Tech. Report No. 320–2993, IBM—Philadelphia Scientific Center (1970).

[34] White, W. W., "Mathematical Programming, Multicommodity Flows, and Communication Nets," Proceedings of the Symposium on Computer—Communications Networks and Teletraffic at Polytechnic Institute of Brooklyn, 325–334 (1972).

[35] Wollmer, R. D., "Multicommodity Networks with Resource Constraints: The Generalized Multicommodity Flow Problem," Networks, *1*, 245–263, (1972).

STOCHASTIC TRANSPORTATION PROBLEMS AND OTHER NEWTORK RELATED CONVEX PROBLEMS

Leon Cooper and Larry J. LeBlanc

Southern Methodist University
Dallas, Texas

ABSTRACT

A class of convex programming problems with network type constraints is addressed and an algorithm for obtaining the optimal solution is described. The *stochastic transportation problem* (minimize shipping costs plus expected holding and shortage costs at demand points subject to limitations on supply) is shown to be amenable to the solution technique presented. Network problems whose objective function is non-separable and network problems with side constraints are also shown to be solvable by the algorithm. Several large stochastic transportation problems with up to 15,000 variables and non-negativity constraints and 50 supply constraints are solved.

0. INTRODUCTION

In previous papers (see [2, 3]), the Frank-Wolfe convex programming algorithm has been applied to the network equilibrium problem and to a nonlinear transportation-production problem. The results indicated that the algorithm is an efficient computational method compared with existing alternatives. In this paper, we point our a much wider class of problems for which this approach has proven to be surprisingly efficient. In particular we give several examples of relatively large-scale (5,000–15,000 variables) nonlinear programming problems with very modest computational times (30–50 seconds on a CDC Cyber 70, model 72 for the 5,000 variable problems).

The class of problems we consider is of the form

$$\begin{aligned} &\text{Min } f(x) \\ &Ax = b \qquad \text{(NLP)} \\ &x \geq 0 \end{aligned}$$

where $x \epsilon R^n$, A is $m \times n$, $b \epsilon R^m$ and $f(x)$ is a convex differentiable function. We consider problems where (NLP) would be readily solvable if $f(x)$ were a linear function. Examples include convex minimum cost flow problems (single or multicommodity), stochastic transportation problems and network problems with linear or convex side constraints. In the latter case the side constraints can be included in the objective function by means of a penalty function, which results in a problem of the form (NLP). An additional class of problems to which this general approach is applicable

are those whose constraints exhibit a network structure and have nonseparable (convex) objective functions.

The Frank-Wolfe algorithm is described in Zangwill [6], pp. 158–162. Briefly the method is as follows: The algorithm determines a search direction by solving the linear programming subproblem of minimizing a first-order Taylor's approximation to $f(\cdot)$ about a feasible solution x^k, subject to the constraints of (NLP):

$$\begin{aligned} \min_z \; & f(x^k)+\nabla f(x^k)\cdot(z-x^k) \\ & Az=b \qquad\qquad (SP) \\ & z\geq 0 \end{aligned}$$

The terms $f(x^k)$ and $\nabla f(x^k)\cdot x^k$ are constant and may be omitted when solving the subproblem (SP). If z^k is the optimal solution to (SP), then the search direction is defined to be:

$$d^k \equiv z^k - x^k$$

After searching in this direction, a new point x^{k+1} is obtained and the process is repeated. The proof that the sequence x^k converges to x^*, the optimal solution to (NLP), is given in [6].

Although the computational effort of solving the linear program (SP) may seem to be unnecessarily high just to find a search direction, this has not proved to be the case. In the problems described in this paper, problem (SP) is solvable by inspection or by simple network techniques. Although the Frank-Wolfe algorithm is known to be only linearly convergent, computational results have indicated that for large scale problems of the type described above, the total computational effort (i.e., the number of iterations multiplied by the computational effort per iteration) is considerably less than that of alternative solution techniques. Several examples are given in Section 5.

At each iteration of the Frank-Wolfe procedure, a lower bound on the optimal value of (NLP) is available by noting that:

$$f(x^*)\geq f(x^k)+\nabla f(x^k)\cdot(x^*-x^k)\geq f(x^k)+\nabla f(x^k)\cdot(z^k-x^k)$$

See [3] for a derivation of this result.

1. THE STOCHASTIC TRANSPORTATION PROBLEM

The stochastic transportation problem is concerned with how to choose quantities to be shipped from supply points to demand points when the requirements at destinations are random variables rather than known constants. Since customer demands are not known, if a certain quantity of material is shipped to some destination, then an *expected* holding cost and and an *expected* shortage cost is incurred. In the stochastic transportation problem, we wish to choose amounts to be shipped from each supply point to each demand point in order to minimize shipping costs (which are deterministic) plus expected holding and shortage costs. The problem is shown to be a convex nonlinear programming problem in [1]. In the stochastic transportation problem, we consider m existing supply points, each with a known supply of S_i, $i=1, 2, \ldots, m$. We are also given n demand points,

each with demand D_j, $j=1, 2, \ldots, n$, where D_j is a *random variable* with density function $\phi_j(v)$. We then wish to choose shipments x_{ij} from supply points to demand points to

$$\text{Min} \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij}x_{ij} + \sum_{j=1}^{n} \left[h_j \int_0^{y_j} (y_j - v)\phi_j(v)dv + p_j \int_{y_j}^{\infty} (v - y_j)\phi_j(v)dv \right] \tag{1}$$

$$y_j = \sum_{i=1}^{m} x_{ij} \qquad j=1, 2, \ldots, n \tag{2}$$

$$\sum_{j=1}^{n} x_{ij} \leq S_i \qquad i=1, 2, \ldots, m \tag{3}$$

(STP)

$$x_{ij} \geq 0 \qquad i=1, 2, \ldots, m;\ j=1, 2, \ldots, n \tag{4}$$

where

c_{ij}=unit shipping cost from supply point i to destination j.

h_j=unit holding cost at destination j.

$\phi_j(v)$=probability density function for demand at destination j.

p_j=unit shortage cost at destination j.

S_i=supply at i.

In (1), y_j is the total amount shipped into destination j from all sources, and so

$$\int_0^{y_j} (y_j - v)\phi_j(v)\,dv$$

is the expected amount of material which must be held at destination j. Similarly,

$$\int_{y_j}^{\infty} (v - y_j)\phi_j(v)\,dv$$

is the expected shortage. Therefore the expected holding cost at destination j is

$$h_j \int_0^{y_j} (y_j - v)\phi_j(v)dv$$

and the expected shortage cost is

$$p_j \int_{y_j}^{\infty} (v - y_j)\phi_j(v)dv$$

Thus we are minimizing shipping costs plus expected holding and shortage costs. Costraints (2) are definitional constraints relating y_j, the total amount shipped into demand point j, to the shipments x_{ij}. Constraints (3) are supply constraints for each supply point i.

We now discuss the computational aspects of the Frank-Wolfe algorithm for (STP) and indicate why it is more efficient than any other technique that we are aware of for this problem. In the Frank-Wolfe algorithm a linear programming subproblem must be solved at each iteration. The objective function of (SP) changes at each iteration; at iteration k the objective functoin is $\nabla f(x^k)$, where x^k is the current vector of flows. It is shown in [1] that the nonlinear objective function (1) can be written

$$f(x) = \sum_{ij} c_{ij}x_{ij} + \sum_{j=1}^{n} \left[h_j y_j + (h_j + p_j) \int_{y_j}^{\infty} (v - y_j)\phi_j(v)dv \right] \tag{5}$$

where x has components x_{ij}. Now define

$$g_j(y_j)=h_jy_j+(h_j+p_j)\int_{y_j}^{\infty}(v-y_j)\phi_j(v)dv \tag{6}$$

An examination of (5) shows that

$$\frac{\partial f(x)}{\partial x_{ij}}=c_{ij}+\frac{\partial g_j(y_j)}{\partial x_{ij}}$$

Now

$$\frac{\partial g_j(y_j)}{\partial x_{ij}}=\frac{\partial g_j(y_j)}{\partial y_j}\cdot\frac{\partial y_j}{\partial x_{ij}}=h_j+(h_j+p_j)\left[-(y_j-y_j)\phi_j(y_j)+\int_{y_j}^{\infty}(0-1)\phi_j(v)dv\right]$$

The first term within the brackets is zero, and we have therefore:

$$\frac{\partial f(x)}{\partial x_{ij}}=c_{ij}+h_j-(h_j+p_j)\int_{y_j}^{\infty}\phi_j(v)dv \tag{7}$$

Now if x^k is a feasible solution for (STP), we define:

$$\left.\frac{\partial f}{\partial x_{ij}}\right|_{x=x^k}\equiv\bar{c}_{ij}$$

Then when using the Frank-Wolfe algorithm on the stochastic transportation problem, the subproblem (SP) becomes

$$\text{Min}\sum_{i=1}^{m}\sum_{j=1}^{n}\bar{c}_{ij}z_{ij}$$

$$z_{ij}\geq 0 \qquad i=1,2,\ldots,m; \qquad j=1,2,\ldots,n \tag{8}$$

$$\sum_{j=1}^{n}z_{ij}\leq S_i \qquad i=1,2,\ldots,m$$

The constraints in the linear program (SP) never change; they are supply constraints and nonnegativity at each iteration. Constraints (2) are not included because they are only definitional; they would not be present in the (STP) and therefore would not be in the subproblems (SP). Instead, the variables y_j would be replaced by $\sum_{i=1}^{m}x_{ij}$ in the objective function (1) and deleted from the problem.

The key to the computational success of the Frank-Wolfe algorithm for solving the stochastic transportation problem is that there are no constraints requiring any material to be shipped to the destinations. Thus, since the objective function is linear in the subproblems, each subproblem decomposes into m separate problems, one for each supply point. Because of this the optimal solution to each subproblem is obvious by inspection. The optimal solution to subproblem (SP) is obtained at each iteration by examining each source i and choosing

$$\bar{c}_{ik}=\min_j\bar{c}_{ij}$$

The optimal solution to each (SP) is then given as follows:

$$\text{If } \bar{c}_{ik}<0, \begin{cases} z^*_{ik}=S_i \\ z^*_{ij}=0 \quad j=1, 2, \ldots, n;\ j\neq k \end{cases} \tag{9}$$

$$\text{If } \bar{c}_{ik}\geq 0,\ z^*_{ij}=0 \qquad j=1, 2, \ldots, n \tag{10}$$

In other words, the optimal solution is to ship everything available to demand point k if $\bar{c}_{ik}<0$, and to ship nothing if $\bar{c}_{ik}\geq 0$. This is possible because there are no constraints requiring material to be shipped to any destination. The simple form shown in (9) and (10) is the reason that the Frank-Wolfe algorithm is so efficient for large stochastic transportation problems. Numerical results are shown in Section 5.

Since the stochastic transportation problem is basically a network problem, an obvious approach is to use piecewise linear approximation and a minimal cost flow algorithm. To compare the Frank-Wolfe technique with piecewise linear approximation, a ten-source, 100 destination problem was solved by both algorithms. Shipping costs, supplies, demands, etc., for this problem were generated as random numbers. The data are described in Section 5.

The linear cost network used to model this problem is shown in Figure 1. Nodes 1–10 are supply points and nodes 11–110 are demand points. The 100 sets of 10 nodes each, namely nodes 111–120, . . ., 1,101–1,110 were used so that a 10-piece linear approximation to the expected holding and shortage costs function at the corresponding destination could be used. Ten linear pieces were necessary to achieve 2% accuracy (the same accuracy was demanded of the Frank-Wolfe algorithm).

Computing time for the Out of Kilter algorithm was 48.5 seconds on the CDC Cyber 70, Model 72; the Frank-Wolfe technique took only 10.7 seconds. In addition, the Out of Kilter technique required more than 69,000 words of memory and the Frank-Wolfe technique required less than 24,000 words. Because of this memory difference, the Out of Kilter algorithm was actually 6.5 times more expensive to run.

More importantly, the largest stochastic transportation problem that the Out of Kilter algorithm could handle was 10×100 (total memory available was only 70,000 words). The Frank-Wolfe technique, because of its smaller memory requirements, could handle problems 15 times as large.

Hadley [1] has proposed that Dantzig-Wolfe decomposition be applied on the piecewise linear approximation to a stochastic transportation problem. However, it is known [5] that Dantzig-Wolfe decomposition usually requires much greater computational efforts than the ordinary simplex method for the same problem.

Another solution technique attempted for (STP) was the penalty approach. A quadratic penalty function and the conjugate gradient algorithm [4] were coded for the same 10×100 stochastic transportation problem. This approach was abandoned when it failed to converge even after several hundred seconds of CPU time.

2. MULTI-COMMODITY NETWORK PROBLEMS

A variation of the classic multi-commodity transshipment problem, in which the arcs are uncapacitated but the objective function penalizes flow exceeding a specified threshold, is easily solved by the Frank-Wolfe algorithm. Letting x_{ij}^s denote the flow of commodity s along arc ij,

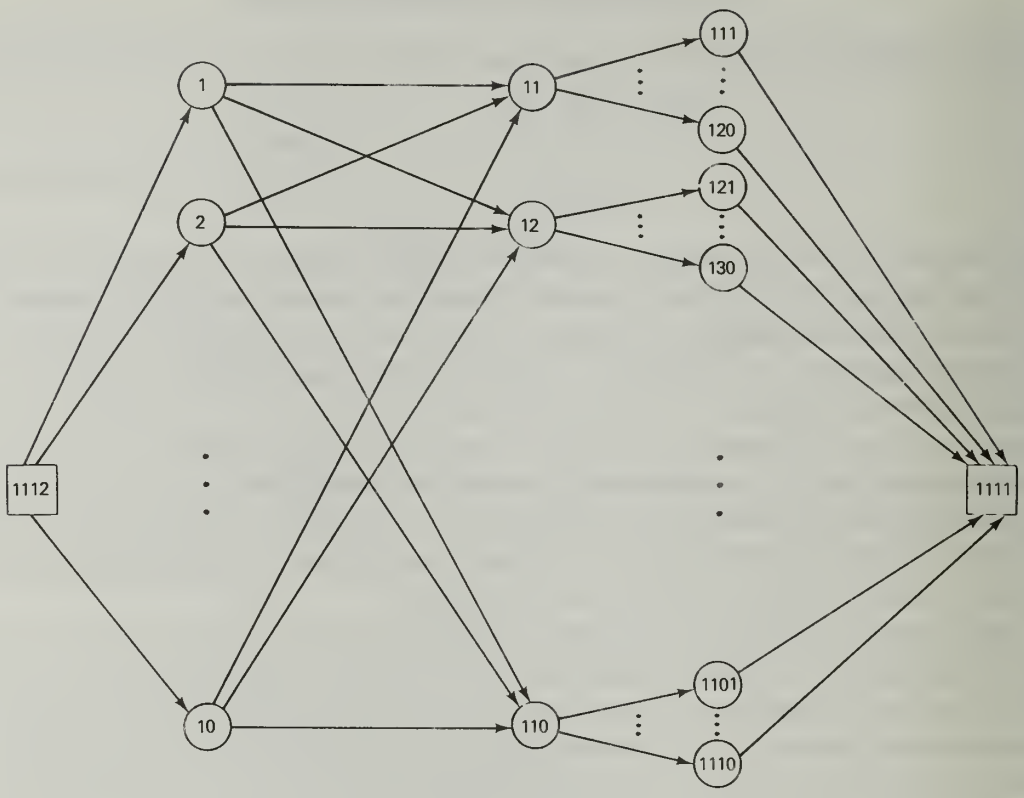


FIGURE 1

S_i^s denote the supply of s at supply point i, and D_j^s denote the demand for s at j, the problem is to

$$\text{Min} \sum_{ij} f_{ij}\left(\sum_{s=1}^{p} x_{ij}^s\right) \tag{11}$$

$$\sum_i x_{ij}^s = D_j^s \qquad \begin{array}{l}\text{all demand points } j=1, 2, \ldots, n \\ \text{all products } s=1, 2, \ldots, p\end{array} \tag{12}$$

$$\sum_j x_{ij}^s \leq S_i^s \qquad \begin{array}{l}\text{all supply points } i=1, 2, \ldots, m \\ \text{all products } s=1, 2, \ldots, p\end{array} \tag{13}$$

$$x_{ij}^s \geq 0 \qquad \text{all } i, j, s \tag{14}$$

In (11) $f_{ij}(\cdot)$ is the shipping cost function for arc ij. In urban network models, $f_{ij}(\cdot)$ is the travel time on arc ij, and x_{ij}^s represents the flow of automobile traffic on arc ij with destination s. The functional form used by the U.S. Bureau of Public Roads is

$$f_{ij}\left(\sum_{s=1}^{p} x_{ij}^s\right) = A_{ij}\left(\sum_{s=1}^{p} x_{ij}^s\right) + B_{ij}\left(\sum_{s=1}^{p} x_{ij}^s\right)^5$$

where A_{ij} and B_{ij} are specified paramenters for arc ij. The paramenter B_{ij} is chosen small (typically 10^{-5}) so that $f_{ij}(\cdot)$ is nearly linear for small flow values. For large flow rates congestion occurs and $f_{ij}(\cdot)$ increases much faster than linearly. When using the Frank-Wolfe algorithm to solve the multi-commodity network problem (11)–(14), each subproblem is a multi-commodity transship-

ment problem with no arc capacities. Because there are no arc capacities, each subproblem splits up into p separate transshipment problems, one for each commodity.

In [2], the network equilibrium problem was solved using the Frank-Wolfe algorithm. The network equilibrium problem is a special type of multicommodity network flow problem in which the constraints specify that a certain amount of automobile traffic must flow between each pair of nodes. As in constraints (12)–(14), there are no arc capacities; instead, the nonlinear objective function prevents excessive flows on any arc at optimality. The Frank-Wolfe subproblems are even simpler for the network equilibrium problem. Since the constraints state only that a certain amount of traffic must travel between each pair of nodes, the Frank-Wolfe subproblems are simply shortest route problems. In [2] a nonlinear program with 1824 variables and non-negativity constraints and 552 conservation of flow constraints was solved in 9 seconds on the CDC 6400 computer. Approximately 20 iterations of the Frank-Wolfe algorithm were required. The same problem was approximated by a piecewise linear function and solved by the simplex method (the multicommodity aspect required that the simplex method be used). Computing time on the 6400 was 11 minutes and 40 seconds—more than 77 times as great as the computation time of the Frank-Wolfe algorithm! The optimal values obtained from Frank-Wolfe algorithm and from the simplex method differed only by 1.2%.

3. NON-SEPARABLE NETWORK PROBLEMS

It is possible that a mathematical programming model may exhibit a network structure in its constraints but have a general (i.e., non-separable) convex objective function. The problem would then be considered a general nonlinear programming problem which would not be readily amenable to the conventional approach of piecewise linear approximation. The problem we consider is again (NLP) where the matrix A has a transportation or other simple structure characteristic of networks. A simple example of such a problem is as follows. Consider the job shop network of Figure 2. Items continuously flow from the source (node 1) to the destination (node 19). Each item must be processed first on any one of the four lathes (nodes 2, 3, 4, or 5), and then on any one of the two drill presses, unless lathe 5 is used. From the figure, we see that the items must then be shaped, sanded, and welded. Finally, each item must be drilled again. It should be emphasized that the drill presses associated with nodes 17 and 18 are identical machines to the drill presses designated by nodes 6 and 7, respectively. We see from the figure that all jobs flowing along arcs (2, 6), (3, 6), and (4, 6) must be processed by the operator of the drill press designated by node 6. In addition, jobs flowing along arcs (12, 17), (13, 17), (14, 17), (15, 17), and (16, 17) must be processed by the drill press operator at node 17. However, since drill presses 6 and 17 are identical and have the same operator, the processing cost associated with these arcs is

$$f(x)=f(x_{26}, x_{36}, x_{46}, x_{12,17}, x_{13,17}, x_{14,17}, x_{15,17}, x_{16,17}) \tag{15}$$
$$=f(x_{26}+x_{36}+x_{46}+x_{12,17}+x_{13,17}+x_{14,17}+x_{15,17}+x_{16,17})$$

We assume that $f(\cdot)$ is convex; a common example would be when overtime costs must be paid or an additional operator hired if the total flow into nodes 6 and 17 is too large. Letting $t(\bar{x})=x_{26}+x_{36}+$

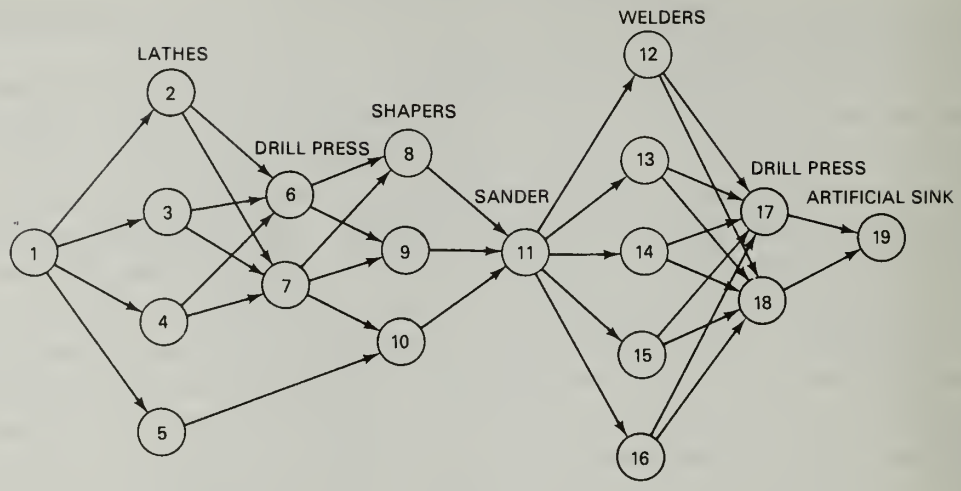


FIGURE 2

$x_{46}+x_{12,17}+x_{13,17}+x_{14,17}+x_{15,17}+x_{16,17}$, we have $f(\bar{x})=f(t(\bar{x}))$. A typical functional form would be that shown in Figure 3:

$$f(t(x))=ct+dt^2 \tag{16}$$

The parameter d is chosen sufficiently small so that $f(\cdot)$ is nearly linear except for flow values exceeding some threshold.

Obviously the function in (16) contains many cross products, and so separable programming cannot be used directly. Although separability can easily be induced in (16), to do so would destroy the network structure of the constraints, leaving a general linear programming problem as an approximation to the scheduling problem.

On the other hand, when the Frank-Wolfe algorithm is used on a network problem with non-separable objective function, each subproblem is obviously a network problem. In fact, if the scheduling problem for the network in Figure 2 is a minimum cost flow problem (with increasing marginal costs instead of capacities), then each subproblem is a shortest path problem.

Many other examples of non-separable problems occur in scheduling of manufacturing processes. Examples include situations where distinct work stations can be monitored by a single individual

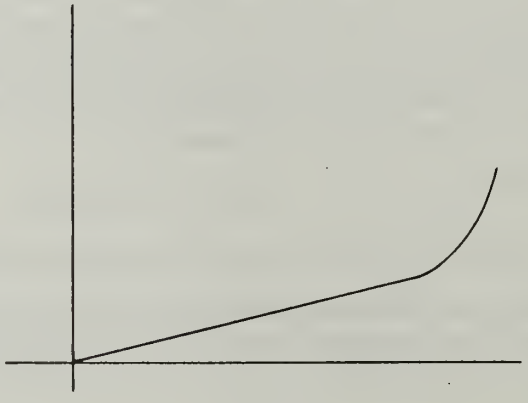


FIGURE 3

at low flow volumes. However, when the total flow volume into these distinct stations becomes large, additional monitors must be procured.

4. NETWORK PROBLEMS WITH SIDE CONSTRAINTS

Frequently network problems which arise in practice are complicated by the presence of a few side constraints. For example, we may have an assignment problem with additional linear or convex constraints which destroy its special structure. Such a problem would be of the form

$$\text{Min}_{x\geq 0}\ c^T x \tag{17}$$

$$Ax=b \tag{18}$$

$$g_i(x)\leq 0 \qquad i=1, 2, \ldots, m \tag{19}$$

where the problem is easily solved in the absence of constraints (19). We can cope with the problem by forming the barrier function [6]:

$$B_k(x)=c^T x-\sum_{i=1}^{m}\frac{r_k}{g_i(x)} \tag{20}$$

We then must solve

$$\text{Min}_{x\geq 0}\ c^T x-\sum_{i=1}^{m}\frac{r_k}{g_i(x)}$$

$$Ax=b$$

This is a convex programming problem amenable to the solution technique described previously, although if there are too many side constraints, the Frank-Wolfe technique would probably have difficulty because of the poor eigenvalue structure of the barrier function. If the objective function (17) were convex instead of linear, the the barrier function (20) would still be convex, and the Frank-Wolfe technique would still be applicable.

5. NUMERICAL RESULTS

To test the efficiency of the Frank-Wolfe algorithm, several large scale stochastic transportation problems were solved. For debugging purposes, a small 3 source, 3 destination problem was used; next fourteen 25 by 200 problems were solved. In all of the stochastic transportaion problems the coordinates of each supply and demand point were chosen as uniform random numbers between 0 and 100. Demand was assumed to be exponentially distributed at each demand point; the parameters λ_j were chosen as uniform random numbers in the interval [.005, .025]. Since the expected demand at destination j equals $1/\lambda_j$, demands were in the range

$$\left[\frac{1}{.025}, \frac{1}{.005}\right]=[40,200].$$

Supplies were chosen randomly in the interval [125,175]; shipping costs were chosen proportional to the distances between supply and demand points. Holding costs were in the range [3, 6], while

shortage costs varied between 20 and 60. This was done so that shortage costs would be significantly greater than holding costs and shipping costs.

In a previous paper [3] which used the Frank-Wolfe algorithm, the authors also studied problems in which the objective function included linear functions and non-linear functions. For such problems, it was noted that the number of iterations of the Frank-Wolfe algorithm required for any given degree of accuracy depended upon the ratio of the non-linear costs to the linear costs. For that reason, in this paper we have chosen costs such that at optimality the nonlinear expected holding and shortage costs accounted for approximately 95–97% of the total costs. This was accomplished by choosing appropriate proportionality constants in calculating the shipping costs.

Computational results for the above problems are as follows. For the 3 x 3 problem the number of iterations for a solution accurate to within 5% of the lower bound was 5; 9 iterations were required for 2% accuracy. For the fourteen 25 x 200 problems, average computing times and numbers of iterations are shown in Table 1. Remarkably, we see that the number of iterations required for 2% accuracy for the 9 variable problem and the 5000 variable problems differed by only a factor of two.

TABLE 1. *Average number of iterations and CPU time (CYBER 70, Model 72)*

	Average number of Iterations and Std. Dev.	Average CPU Time (Seconds)
5% accuracy------	10. 2±1. 1	28. 5
3% accuracy------	13. 6±1. 6	38. 0
2%----------	17. 4±1. 6	48. 6

Finally, ten 50×300 stochastic transportation problems were solved. These 15,000 variable problems proved more difficult to solve as accurately as the smaller problems (perhaps because of round off errors). Because of the higher computing times, these latter problems were solved only to 3.5% accuracy. In practical problems as large as these, we feel that 3.5% accuracy is probably more accurate than the values of the parameters used and the assumptions of linear shipping cost and unit holding and shortage costs. Average number of iterations and computer time were 78.9 and 9 minutes, 55 seconds, respectively.

It appears from the above numerical results that large-scale stochastic transportation problems can be solved quite efficiently using the technique described in this paper. These results indicate that the number of iterations increases very slowly with problem size. Also, the computational effort for each iteration consists of scanning each column of an m × n matrix exactly once and a one dimensional search of mn variables. Therefore the computational effort for each iteration increases only linearly with problem size.

6. CONCLUSION

We have addressed a class of convex network problems and have shown that, by capitalizing on their structure, the Frank-Wolfe algorithm becomes extremely efficient for large-scale problems. Several different examples of convex network problems have been considered. In each case, we have

shown that the structure of the problem can be exploited to yield an efficient solution algorithm even for realistically large problems.

REFERENCES

[1] Hadley, G., *Nonlinear and Dynamic Programming* (Addison-Wesley, Reading, 1964).

[2] LeBlanc, L. J., E. K. Morlok and W. P. Pierskalla, "An Efficient Approach to Solving the Road Network Equilibrium Traffic Assignment Problem," Transportation Research, *9*, 309–318 (1975).

[3] LeBlanc, L. J. and L. Cooper, "The Transportation-Production Problem," Transportation Science, *8* (4) 344–354 (1974).

[4] Luenberger, D., *Introduction to Linear and Nonlinear Programming* (Addison-Wesley, Reading, 1973).

[5] Wacker, W. D., "A Study of the Decomposition Algorithm for Linear Programming," M.S. Thesis, Washington University (1967).

[6] Zangwill, W., *Nonlinear Programming: A Unified Approach* (Prentice-Hall, Englewood Cliffs, 1969).

PROBABILISTIC IDEAS IN HEURISTIC ALGORITHMS FOR SOLVING SOME SCHEDULING PROBLEMS

I. Gertsbach

Ben Gurion University of the Negev,
Beersheva, Israel

ABSTRACT

Heuristic algorithms for positioning a maximal number of independent units and constructing a schedule with minimal fleet size are proposed. These algorithms consist of two stages: defining the "leading" job and finding an optimal position for it. Decisions on both stages use some special criteria which have a probabilistic interpretation. Some experimental data are given.

1. INTRODUCTION

The solution of many scheduling problems can be reduced to positioning, under certain conditions, units and zeroes in matrices. For example, in production schedules, the *k-th* element of *j-th* row, a_{jk}, may represent the *k-th* time interval for the *j-th* machine; $a_{jk}=1$ (0) means that in this time interval the machine is working (vacant). Constructing a schedule for processing a group of parts on a group of machines means to find "good" positions in such a way that the prescribed technological sequence is guaranteed for every part, and every machine is functioning under proper conditions. Another example is that of transportation schedules, where the *j-th* row corresponds to a discrete time scale for the *j-th* station, units represent busy time moments, and zeroes represent non-occupied moments. A trip is defined by a sequence of stations and intervals between departures and arrivals. Selecting a trip in a schedule means, in fact, a displacement of "chains" consisting of units in station time scales. Such displacement, of course, must be done while satisfying given safety conditions.

Many deterministic procedures have been proposed for solving problems similar to those mentioned above. They are based on graph theory [1], discrete linear programming [6] or heuristic approaches. For solving certain specific problems, many effective algorithms have been proposed, e.g., for Johnson's problem [5].

If we try to formalize a real scheduling situation, we have to take into consideration many restrictions and the essential increase of the size of the problem that will make our "exact" solution more realistic and applicable but at the same time much more difficult to find.

We can imagine the process of schedule construction as a sequential procedure. At each step, we solve a subproblem for disposing a specific group of units (variables) in free positions. This must be done, of course, by preserving all prescribed requirements. When the size of the matrix is large (in a realistic schedule it might be 200×200) the positioning of units fixed earlier might be assumed as random. This suggests the idea of approaching scheduling from a probabilistic point of view.

This paper contains several heuristic algorithms which use probabilistic ideas. Some formal statements are also proved which serve as a guide for constructing heuristic algorithms. These algorithms might be useful in constructing large schedules where the traditional methods are not applicable.

Section 2 deals with the problem of positioning independent and dependent units in a matrix with a given set of free and forbidden positions in every row. As a guide for proposed algorithms, two statements will serve. The first asserts that a specially chosen measure (entropy) is closely connected with a possibility of disposing the maximal number of independent units. The second asserts that there is a "universal" row which has all free positions available. By combining these statements, an algorithm is proposed and some numerical results are given. Section 3 deals with a job-shop scheduling problem in which each job occupies a machine for a fixed time length but its starting time can be shifted within certain tolerances. The problem is to find actual starting times for each job providing a minimal number of busy machines (minimal fleet size). A heuristic algorithm similar to the previous one is given and some experimental data are presented. Section 4 contains some concluding remarks about probabilistic interpretation of the sequential procedure of constructing large-size schedules.

2. POSITIONING OF UNITS

2.1 Basic Definitions

A square matrix $A=||a_{jk}||_{M\times M}$ is given. I_j denotes the set of all free positions in the j-*th* row. I_j is nonempty, $j=1, \ldots, M$. Other positions in j-*th* row are forbidden. The set of units disposed in the free positions is called independent (ISU) if in every row and column there is not more than one unit. The deterministic algorithm for finding a maximal ISU is well known [1]. We give an alternative heuristic algorithm which exploits some probabilistic ideas.

Let $\tilde{A}=||\tilde{a}_{jk}||_{M\times M}$ be a matrix with elements

$$\tilde{a}_{jk}=\begin{cases}0, \text{ if } a_{jk}\notin I_j;\\ |I_j|^{-1}, \text{ if } a_{jk}\in I_j.\end{cases} \tag{2.1}$$

($|I|$ denotes the number of elements in the set I).

The values

$$b_k=M^{-1}\sum_{j=1}^{M}\tilde{a}_{jk},\ k=\overline{1, M} \tag{2.2}$$

form a discrete M-point distribution. Let us introduce its entropy

$$E_A=-\sum_{k=1}^{M} b_k \log b_k \tag{2.3}$$

and notice that its maximal value $E^*=\log M$ corresponds to a uniform distribution $b_k=M^{-1}$.

2.2 Kordonsky's Statement

The following statement was asserted by H. Kordonsky (1972).

STATEMENT 1: If $E_A=E_A^*$ then there is an *ISU* which contains exactly M elements.

PROOF: The condition $E=E^*$ means that $\tilde{A}$ is bistochastic. Applying the Birkhoff theorem (see [7] Part 2, Section 1.4–1.7) one obtains $\tilde{A}=\sum_{i=1}^{M} C_i\tilde{A}_i^*$, $C_i\geq 0$, $\sum_{i=1}^{M} C_i=1$, where each $\tilde{A}_i^*$ is a 0/1 matrix with an ISU containing M elements. This completes the proof.

Now let $A=||a_{jk}||_{MxN}, M<N$, be a rectangular matrix. We construct in a similar way the matrix $\tilde{A}=||\tilde{a}_{jk}||_{MxN}$ and define the distribution $\{b_k, k=\overline{1, N}\}$. The entropy of it is maximal when all $b_k=\frac{1}{N}$.

STATEMENT 2: If $b_k=N^{-1}$ then there is an ISU for $A=||a_{jk}||_{MxN}$, $M<N$, which contains M elements.

PROOF: Assume that in A it is possible to position only $R<M$ units. Without loss of generality let $a_{jj}=1$, $j=\overline{1, R}$ (see Figure 1). According to the hypothesis, the submatrix E has no free positions. $b_k=1/N$ means that in every column of D there must be free positions. Again let

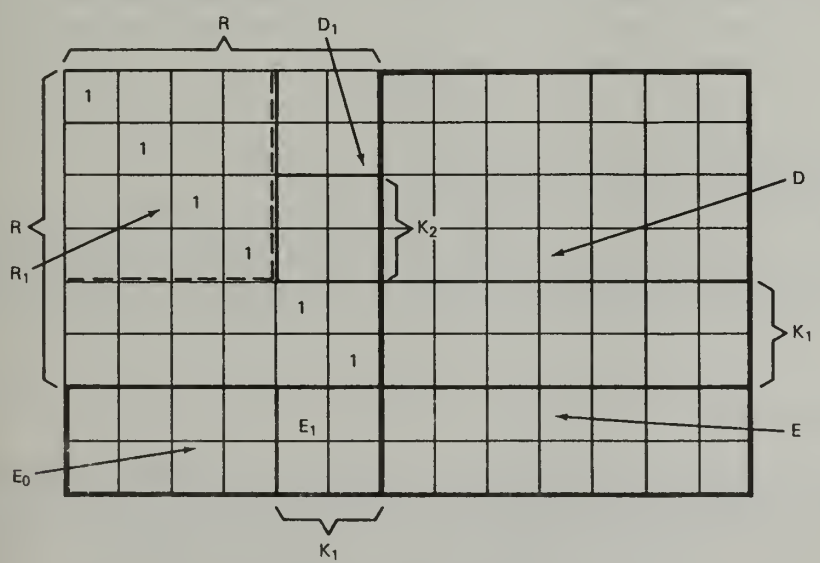


FIGURE 1. Decomposition of the matrix A

us assume that these positions all are in the submatrix consisting of k_1 lowest rows of D. Then the submatrix E_1 (see Figure 1) has no free positions. If it were not true, one could place a unit in that free position, shift a corresponding unit from R to D and in such a way increase the number R. Now consider the submatrix $R_1=||a_{jk}||$, $j, k=\overline{1, (R-k_1)}$ and the corresponding matrix $\tilde{A}_1=||\tilde{a}_{jk}||$, $j, k=\overline{1, (R-k_1)}$. According to (2.2) and the equality $b_k=1/N$, $M^{-1}\sum_{j=1}^{R-k_1}\tilde{a}_{jk}\leq b_k=N^{-1}$, or $\sum_{j=1}^{R-k_1}\tilde{a}_{jk}\leq M/N<1$.

This means that $\tilde{A}_1$ cannot be stochastic, and according to the definition of R, R_1, and D there must be free elements in the first $(R-k_1)$ rows placed in D_1 (see Figure 1). Suppose again that they occupy the last k_2 rows. Repeating the above reasoning we shall conclude that E_o does not contain free elements at all. This contradicts the assumption that I_j, $j=1, \ldots, M$, are nonempty.

2.3 Universal rows

Now let us assume that for a given $A=||a_{jk}||_{MxM}$ the ISU consists of M elements. The element a_{jk} is called admissible if after cancelling the *j-th* row and *k-th* column, the ISU for remaining part of A consists of $(M-1)$ elements. If for all $k \in I_j$, a_{jk} is admissible, the *j-th* row is called a universal row (U-row).

STATEMENT 3: If ISU has M elements then A has a U-row.

PROOF: Let us consider one ISU in A. It is clear that it is worthwhile to investigate only the case when $|I_j|\geq 2$. $j=1, \ldots, M$. We shall call a cycle a closed polygon consisting of alternate horizontal and vertical segments. It is easy to verify that if $|I_j|\geq 2$ it is always possible to find a cycle in which free positions and units alternate. Now, in this cycle, the vertex adjoining some unit is also an admissible position for it. Let us cancel all nonoccupied elements in a constructed cycle and repeat the procedure. Sooner or later we shall get a matrix with one or more rows containing only one noncancelled position. Every such row is, according to the definition, a U-row.

REMARK: If *j-th* row is a U-row, $|I_j|\geq 2$ then from the proof it follows that in the *k-th* column, $k \in I_j$, there must be at least one free position a_{rk}, $r\neq j$.

The existence of a U-row is very important for an algorithm having as its purpose to position a maximal number of independent units. If we could "identify" the U-row we could safely start with positioning a unit in it. A heuristic procedure for identifying the U-row is proposed in Section 2.5.

2.4 Random Matrices

The object of consideration in Sections 2.4–2.5 will be square matrices which have a random arrangement of free and forbidden elements. Let μ_j be integers, $0\leq\mu_j\leq M-1$, $j=\overline{1, M}$ and A_o is a matrix with fixed ISU—set of size M. Consider a class of all matrices $\mathcal{A}_{A_o}(\mu_1, \ldots, \mu_M)$ which contains all matrices obtained from A_o by means of all possible random choice of μ_j free positions among $(M-1)$ in the *j-th* row. We assume that all permutations have the same probability. Thus we can speak about a probability of an event "*j-th* row is a U-row." This probability is denoted by $P_j(\mu_1, \ldots, \mu_M)$. Let $j_2>j_1$, $\mu_{j_1}>\mu_{j_2}$. Then we have the following:

STATEMENT 4:

$$P_{j_1}(\mu_1, \ldots, \mu_{j_1}, \ldots, \mu_{j_2}, \ldots, \mu_M) \leq P_{j_2}(\mu_1, \ldots, \mu_{j_1}, \ldots, \mu_{j_2}, \ldots, \mu_M) \tag{2.4}$$

PROOF: The proof follows from the relationships:

$$P_{j_1}(\mu_1 \ldots \overset{\downarrow j_1}{\mu_{j_1}} \ldots \overset{\downarrow j_2}{\mu_{j_2}} \ldots \mu_M) = P_{j_2}(\mu_1 \ldots \overset{\downarrow j_1}{\mu_{j_2}} \ldots \overset{\downarrow j_2}{\mu_{j_1}} \ldots \mu_M)$$

$$\leq P_{j_2}(\mu_1 \ldots \overset{\downarrow j_1}{\mu_{j_2}} \ldots \overset{\downarrow j_2}{\mu_{j_2}} \ldots \mu_M) \leq P_{j_2}(\mu_1 \ldots \overset{\downarrow j_1}{\mu_{j_1}} \ldots \overset{\downarrow j_2}{\mu_{j_2}} \ldots \mu_M) \tag{2.5}$$

The first equality follows from symmetry, the next inequality is true because an addition of free elements to the j_2-*th* row cannot increase the probability that this row would remain a U-row. The last inequality is valid because by adding (j_2-j_1) free elements to j_1-*th* row, the j_2-*th* row remains a U-row.

Statement 4 gives a background for an intuitively clear fact that in a "random" matrix the shortest rows have a larger probability of being U-rows.

2.5 Identification of the *U*-row and Algorithm 1

Assume that one unit is placed in every row of A in such a manner that a certain position in a j-*th* row is chosen with probability $|I_j|^{-1}$, $j=\overline{1, M}$. Let us consider the j_o-*th* row. The element $a_{j_o k}$ is called marked if there is at least one unit in k-*th* column, with $j \neq j_o$. The j_o-*th* row is called occupied if all its free elements are marked. Now a long sequence of random placements of units is performed. For each member of that sequence the event "j_o-*th* row is occupied" or "j_o-*th* row is nonoccupied" is observed. After completing this sequence, we estimate the frequency P_j^* for the event "the j-*th* row is occupied." The heuristic principle for identification of U-row is the following. The row j_o is declared to be a U-row if

$$P_{j_o}^* = \max_{1 \leq j \leq M} P_j^* \tag{2.6}$$

The reasons for this principle might be Statement 4 and the following remark. Among all rows, the j_0-*th* row determined by (2.6) is the most "vulnerable;" it has more chances than other rows to be lost (occupied) by placing units in other rows.

Now we formulate a rule for finding a suitable position in a given row (assumed to be a U-row).

The identification of a U-row might not always be correct. Therefore it would be desirable to supplement it by an optimal placement of the unit in that row. Statements 1 and 2 assert that the maximal number in an ISU is guaranteed when the entropy E_A is maximal. Then, it seems to be expedient to maximize the entropy by placing the unit in a given row. The latter might be performed n the following manner. Assume that a unit must be placed in the j-*th* row. The element a_{jk}, $k \epsilon I_j$ s chosen and put $\tilde{a}_{jk}=1$. For other rows, the values $\tilde{a}_{jk}$ are defined according to (2.1), and then the value of the entropy $E_A(j, k)$ is calculated. The decision $a_{jk_0}=1$ is made, if

$$E_A(j, k_o) = \max_{k \epsilon I_j} E_A(j, k) \tag{2.7}$$

Combining the heuristic principle for identification of a U-row and for placing a unit in it, we obtain the following:

ALGORITHM 1:

a. A random placement of units is repeated K times and the U-row j_0 is found (see (2.6)).
b. In j_0-*th* row a unit is placed according to (2.7) on a certain place $a_{j_0k^*}$.
c. The j_0-*th* row and k^*-*th* column are cancelled and the process continues for the rest of the matrix A

2.6 Some Experimental Data

By help of a computer, units were positioned in 50 matrices of size 40×40.

Every row contained from 1 to 5 free elements and the matrices were constructed in such a manner that the ISU contained 40 elements. Algorithm 1 positioned 1997 units and, therefore, failed in positioning three units.

2.7 Nonindependent Units

Algorithm 1 has two important features: it is sequential and has two stages. That means that, first, a "leading" row is chosen and, second, a position for the unit in this row is sought. The algorithm requires only a "forward pass" and can be performed quickly even when it is applied to a large-size matrix. The similar algorithm might be constructed for a much more difficult problem, for example, for the problem of disposition of "chains" consisting of units. Such problems arise in aircraft scheduling, where a_{jk} means that the k-*th* time period in the j-*th* airport is occupied; the chain of units corresponds to a trip (string). In this case the units might be called "dependent" because between the units in one row belonging to different trips, there must be a safety delay time. The realistic criterion for a schedule would be the maximal number of trips "packed" into a given matrix. References [3, 4] contain some facts about the realization of a sequential two-stage algorithm for a large-size aircraft schedule.

2.8 Shortening the Sequential Procedure

In computerized experiments it was stated that if on a certain stage of scheduling performed by Algorithm 1, the arrangement of rows was done according to the decrease of P_j^* (see (2.6)), then without big changes the same arrangement remains valid during several steps of the procedure. Thus, by composing a schedule for 1500 trips [4], it is enough to calculate the priorities only 5–6 times after finding positions for every 250–300 trips.

3. POSITIONING OF SEGMENTS

3.1 The Statement of the Problem

We have M jobs (segments): the j-*th* one has a length of l_j units of time and a starting time t_j which has to be within given limits a_j, b_j: $a_j \leqslant t_j \leqslant b_j$. Every job must be performed without intermissions. Then there is a set of machines; each machine can perform every job and only one job can be performed in a given time on a given machine.

The problem is to find such values t_j, $j=\overline{1, M}$, so that the number of needed machines (fleet size) is minimal. We assume that a_j, b_j, t_j, l_j are integers; rows in a matrix A represent jobs and columns represent time intervals. If the *j-th* job is performed with a starting time t_j, we put $a_{jk}=0$ for $k<t_j$, $k>t_j+l_j-1$ and $a_{jk}=1$ otherwise. In that case the number of machines in the fleet is given by a value $H=\max_k \sum_{j=1}^{M} a_{jk}$ (See [1], Chapter 2, Section 9, Special case of Dilworth's theorem).

3.2 Local Criterion

In the case when the starting times are given only within their ranges $[a_j, b_j]$, we propose a new scalar quantity which is in some sense similar to the value H and gives a characterization of the mutual disposition of all jobs.

Suppose that the *j-th* job can have its starting time equal to one of values $x\epsilon(a_j, b_j)$ with equal probabilities $(b_j-a_j+1)^{-1}$. The matrix $\tilde{A}=||\tilde{a}_{jk}||_{M\times N}$, $N=\max_j (b_j+l_j-1)$, is defined in a following way. $\tilde{a}_{jk}$ is equal to the probability that the position a_{jk} would be busy when the starting time of *j-th* job is randomly chosen within its range. Let $b_k=\sum_{j=1}^{M} \tilde{a}_{jk}$, $k=\overline{1, N}$. We now introduce the entropy of distribution $\{b_k^o\}_1^N$ obtained from $\{b_k\}$ after norming:

$$E_N=-\sum_{k=1}^{N} b_k^o \log b_k^o \tag{3.1}$$

This value supplies us with information about the possible minimal number of machines in the fleet. Thus the small value of E_N will occur when many jobs are concentrated within the same time period. That would lead to a large number of busy machines.

Now we give a local criterion for choosing a position for a given j-th job.

Let the starting time t_{j_1} lie between a_j and b_j: $a_j\leq t_{j_1}\leq b_j$. The value t_{j_1} is chosen and the values $\tilde{a}_{jk}$ are defined in the following way: $\tilde{a}_{jk}=1$ if $k=t_{j_1}, t_{j_1}+1, \ldots, t_{j_1}+l_j-1$ and $\tilde{a}_{jk}=0$ otherwise. Now the value of $E(j_1)$ is calculated according to (3.1). The time t_{j^*} is assumed "the best" if

$$E(j^*)\geq E(j_1) \text{ for all } t_{j_1}\epsilon[a_j, b_j] \tag{3.2}$$

3.3 Determination of Priority

Let us assume that the exact starting times for jobs with numbers $j\epsilon S^+$ are already chosen. Every job j, $j\notin S^+$ is randomly placed within its range. A number j_o, $j_o\notin S^+$ is fixed and the difference $\Delta H(j_o)$ is calculated:

$$\Delta H(j_o)=\max_{1\leq k\leq N} \sum_{j=1}^{M} \tilde{a}_{jk}-\max_{1\leq k\leq N} \sum_{j=1}^{M} \tilde{a}_{jk}(1-\delta_{jj_o}). \tag{3.3}$$

δ_{jj_0} is a Kroneker symbol. The value $\Delta H(j_o)$ has the following meaning: it is the additional number of machines needed for j_o-*th* job when this job will be added to other jobs. The procedure of random assignment of jobs is repeated K times and the average value of $\overline{\Delta H(j_o)}$ is calculated for every job $j_o\notin S^+$. We shall give the highest priority to the job number j^* which satisfies the following inequality:

$$\overline{\Delta H}(j^*)\geq\overline{\Delta H}(j_o),\ j_o\epsilon S^+ \tag{3.4}$$

ALGORITHM 2:

a. According to the description given in section 3.3, the job j_o with highest priority is found.
b. According to local criterion, (see section 3.2) the optimal position for the job j_o is found in the j_o-*th* row.
c. The job j_o is added to the set S^+ of scheduled job and then phase a is repeated.

The heuristic justifying of this algorithm is as follows. The highest priorities will be given to the most "difficult" jobs wh'ch demand for themselves an additional machine with higher probability. The best position for a given job is the position which provides maximal uniformity according to the local criterion based on entropy. The algorithm is also of a "forward pass" type.

3.4 Some Experimental Data

A series of schedules were constructed on the computer according to Algorithm 2. They demonstrated that this algorithm either provides an optimal solution or gives a schedule having a number of machines in the fleet that has one machine more than in an optimal solution. A typical example for a schedule with 32 jobs in the matrix 32×16 is presented in Table I. The Algorithm 2 found an optimal solution given in this table which required seven machines.

It is interesting to compare this result with three alternative approaches: the schedule is performed absolutely randomly (the priority and the position for every job is chosen randomly); the priority is random, the local criterion corresponds to Algorithm 2; the priority was found according to Algorithm 2, and the position for the job was chosen randomly within its ranges. The results were the following:

(1) The best random schedule among 10,000 done on the computer had 11 machines in the fleet and there were only forty such schedules found.

(2) The random priority in the presence of local optimization gave a schedule requiring eight machines.

(3) The random local policy combined with the priority rule given in the first part of Algorithm 2 gave a schedule requiring 13 machines.

4. SOME CONCLUDING REMARKS

In constructing large-size schedules, the most important and realistic criterion would be a maximal number of jobs performed by a given set of machines and within a given period of time in the presence of some special restrictions. If an algorithm similar to one described in Section 3 is used, then it can happen that at some stage of scheduling a position cannot be found for a certain job. Since the algorithm uses nondeterministic decisions, it seems expedient to estimate its "quality" using the mathematical expectation of the number of jobs disposed in the schedule. That could be done, for example, in the terms of dependent trials when a success in an ordinary trial depends on the number of successful preceding trials. An attempt to evaluate this approach was made in [2].

TABLE 1. *The schedule for 32 jobs.*

The number of job \ Time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	▬	▬							+	+	+	+	+	+	+	+
2	+	+	+	+	+	+					▬	▬	▬	▬	▬	▬
3	+	▬	▬	▬	▬	▬					+	+	+	+	+	+
4	+	+	▬	▬	▬	▬						+	+	+	+	+
5	+	+					▬	▬	▬					+	+	+
6	▬	▬	▬	▬	▬						+	+	+	+	+	+
7	+	+	+	+			▬	▬							+	+
8	+	+	+	+	+	+	+						▬	▬	▬	▬
9	+	+	+	+	+	+				▬	▬	▬	▬			
10	+	+	+	+	+	+	+			▬					+	+
11	▬	▬														
12	+	+	+	+	+				▬	▬	▬	▬				
13	+					▬	▬	▬	▬	+	+	+	+	+	+	+
14			▬	▬	▬						+	+	+	+	+	+
15	+	+					▬	▬	▬	▬				+	+	+
16	+				▬	▬			+	+	+	+	+	+	+	+
17	+	+	+	+								▬	▬	▬	▬	▬
18	▬							+	+	+	+	+	+	+	+	+
19	▬	▬	▬				+	+	+	+	+	+	+	+	+	+
20	+			▬					+	+	+	+	+	+	+	+
21											▬	▬	▬	▬	▬	▬
22	▬	▬	▬													
23	+	+	+	+	+	▬	▬							+	+	+
24	+	+	+	+	+	+	+						▬	▬	▬	▬
25	+	+	+	+	+	+	▬	▬	▬	▬	▬					
26	▬	▬	▬	▬	▬	▬	▬	▬	▬							
27				▬		+	+	+	+	+	+	+	+	+	+	+
28	+	+	+	+	+	+	+	+				▬	▬	▬	▬	▬
29															▬	▬
30	+									▬	▬			+	+	+
31	+				▬	▬			+	+	+	+	+	+	+	+
32	+	+	+	+				▬	▬	▬	▬	+	+	+	+	+
H_k	7	7	7	7	7	7	7	7	7	7	7	6	7	6	7	7

H_k—total number of machines occupied in the *k-th* time period.
+—time not available for the job.
▬—job performed.

REFERENCES

[1] Ford, L. R. and D. R. Fulkerson, *Flows in Networks* (Princeton University Press, Princeton, New Jersey, 1962).

[2] Gertsbach, I. "On a Problem of Choosing the Order of Dependent Trials, Theory of Probability and its Applications," *17*, 709–712 (1972).

[3] Gertsbach, I., M. Maksim, et al., "Heuristic Method for Constructing the Aviation Schedule," In the Coll. Automation in Machinery, Academy of Sciences, (Moscow, 1969) (in Russian).

[4] Gertsbach, I., V. Venevcev, et al., "Central Avia-schedule as a Part of Air Traffic Control System," Proceedings of the First International Traffic Control Session, Sec. 6, 5–26 (Versailles, 1970).

[5] Johnson, S., "Optimal Two and Three-Stage Production Schedules with Setup Times Included," Naval Research Logistics Quarterly, *1*, 61–68 (1954).

[6] Korbut, A. and Yu. Finkelstein, *Discrete Programming* (Nauka, Moscow, 1969) (in Russian).

[7] Marcus, M. and H. Minc, *A Survey of Matrix Theory and Matrix Inequalities* (Allyn and Bacon, Boston, 1964).

FORCE-ANNIHILATION CONDITIONS FOR VARIABLE-COEFFICIENT LANCHESTER-TYPE EQUATIONS OF MODERN WARFARE*

James G. Taylor and Craig Comstock

Naval Postgraduate School
Monterey, California

ABSTRACT

This paper develops a mathematical theory for predicting force annihilation from initial conditions without explicitly computing force-level trajectories for deterministic Lanchester-type "square-law" attrition equations for combat between two homogeneous forces with temporal variations in fire effectivenesses (as expressed by the Lanchester attrition-rate coefficients). It introduces a canonical auxiliary parity-condition problem for the determination of a single parity-condition parameter ("the enemy force equivalent of a friendly force of unit strength") and new exponential-like general Lanchester functions. Prediction of force annihilation within a fixed finite time would involve the use of tabulations of the quotient of two Lanchester functions. These force-annihilation results provide further information on the mathematical properties of hyperbolic-like general Lanchester functions: in particular, the parity-condition parameter is related to the range of the quotient of two such hyperbolic-like general Lanchester functions. Different parity-condition parameter results and different new exponential-like general Lanchester functions arise from different mathematical forms for the attrition-rate coefficients. This theory is applied to general power attrition-rate coefficients: exact force-annihilation results are obtained when the so-called offset parameter is equal to zero; while upper and lower bounds for the parity-condition parameter are obtained when the offset parameter is positive.

1. INTRODUCTION

Deterministic Lanchester-type equations of warfare (see Taylor and Brown [25], Weiss [27]) play an important role in military operations research for developing insights into the dynamics of combat (see, for example, Bonder and Farrell [4], or Bonder and Honig [5]), even though combat between two opposing military forces is a far more complex random process. The classic Lanchester theory of combat (see Dolansky [9]) considered constant attrition-rate coefficients. New operations research techniques for forecasting temporal variations in fire effectiveness (caused by, for example, changes in force separation, combatant postures, target acquisition rates, firing rates, etc.) have generated interest in variable-coefficient combat formulations. Unfortunately, the resultant differential equations are not well studied.

*This research was supported by the Office of Naval Research as part of the Foundation Research Program at the Naval Postgraduate School.

In this paper we present a mathematical theory for predicting battle outcome from initial conditions without explicitly computing force-level trajectories for variable coefficient Lanchester-type equations of modern warfare for combat between two homogeneous forces†. The determination of conditions on initial values that predict force annihilation (in the sense of necessary and/or sufficient conditions) in such Lanchester-type combat leads to some new mathematical problems in the theory of ordinary differential equations. This force annihilation problem may be viewed as either a problem of determining the asymptotic behavior of the solution (depending on given initial conditions) or a problem of determining the range of the quotient of two linearly independent solutions to, for example, the X force-level equation [25]‡. In either case, the classic ordinary differential equation theories (see, for example, Hille [10], Ince [11], and Olver [17]) are inadequate to supply all the answers sought. We show that questions of force annihilation can be reduced to the study of certain "exponential-like" Lanchester functions and may be simply answered by examining certain inequalities involving the initial conditions and possibly by consulting tabulations of new special functions that are suggested here. Our general results apply to a wide class of attrition-rate coefficients (namely, those that yield continuous force-level trajectories).

Thus, in this paper we provide a general theoretical framework for determining force annihilation without explicitly computing force-level trajectories for variable-coefficient Lanchester-type equations of modern warfare. Other modes of battle termination are briefly discussed. We introduce a canonical auxiliary parity-condition problem for such determinations. New exponential-like general Lanchester functions arise from the solution to this problem, and tabulations of these would facilitate force-annihilation prediction. Different mathematical forms for attrition-rate coefficients lead to different auxiliary parity-condition problems. Our theory is applied to general power attrition-rate coefficients· exact force-annihilation results are obtained for cases of "no offset" (modelling, for example, weapon systems with the same maximum effective range); and although qualitative results are obtained, future computational work is required for quantitative results for cases of "offset" (modelling, for example, weapon systems with different maximum effective ranges).

2. LANCHESTER'S CLASSIC FORMULATION

F. W. Lanchester [13] hypothesized in 1914 that combat between two military forces could be modelled by

$$dx/dt=-ay, \qquad dy/dt=-bx, \tag{1}$$

†Bonder and Honig [5] point out, however, that force annihilation may not always be the best criterion for evaluating military operations. See pp. 192–242 of Bonder and Farrell [4] for a detailed Lanchester-type analysis of an attack scenario for which other "end of battle conditions" play the major role in the evaluation process. Nevertheless, it is of interest to be able to easily predict the occurrence of force annihilation. Such results are not only at intrinsic interest but also are useful in the optimization of combat dynamics (see Taylor [21, 23]).

‡Previous work by Bonder and Farrell [4], Taylor [22], and Taylor and Brown [25] shows that new transcendental functions arise even in the case of linear attrition-rate coefficients reflecting weapon systems with different effective ranges (i.e. the coefficients (12) with $\mu=\nu=1$ and $A>0$). For example, the differential equation (52) could not be found among the 445 linear second order equations tabulated in Kamke [12]. Moreover, even when one can express a solution in terms of previously known transcendents, the appropriate tabulations (see, for example, Abramowitz and Stegun [1]) may not exist (see Section 5 of [25]). As the equations of mathematical physics have provided interest in many previously studied transcendents, variable-coefficient Lanchester-type equations provide interest in new transcendents.

with initial conditions

$$x(t=0)=x_0, \qquad y(t=0)=y_0, \tag{2}$$

where $t=0$ denotes the time at which the battle begins, $x(t)$ and $y(t)$ denote the numbers of X and Y at time t, and a and b are nonnegative constants which are today called Lanchester attrition-rate coefficients and represent each side's fire effectiveness. Lanchester (see McCloskey [16] for his influence on operations research) considered this model (1) in order to provide insight into the dynamics of combat under "modern conditions" and justify the principle of concentration.† We will accordingly refer to (1) as *Lanchester's equations of modern warfare*. Various sets of physical circumstances have been hypothesized to yield them: for example, (a) both sides use aimed fire and target acquisition times are constant (see Weiss [27]), or (b) both sides use area fire and a constant density defense (see Brackney [6]).

From (1) Lanchester deduced his classic square law

$$b(x_0^2-x^2(t))=a(y_0^2-y^2(t)). \tag{3}$$

Consider now a battle terminated by either force level reaching a given "breakpoint":‡ for example, Y wins when $x_f=x(t_f)=x_{BP}$ but $y_f>y_{BP}$, where t_f, x_f, y_f denote final values and x_{BP} denotes X's breakpoint. Let us express x_{BP} as a fraction of X's initial strength $f_X{}^{BP}$, e.i. $x_{BP}=f_X{}^{BP}x_0$, and similarly for y_{BP}. It follows from (3) that

$$Y \text{ will win} \longleftrightarrow \frac{x_0}{y_0}<\sqrt{\frac{a\{1-(f_Y{}^{BP})^2\}}{b\{1-(f_X{}^{BP})^2\}}}. \tag{4}$$

We will refer to a battle that continues until the annihilation of one side or the other (i.e. $x_{BP}=y_{BP}=0$) as a *fight-to-the-finish*. In this case (4) becomes

$$Y \text{ will win fight-to-the-finish} \longleftrightarrow \frac{x_0}{y_0}<\sqrt{\frac{a}{b}}. \tag{5}$$

Unfortunately, no relationship similar to (3) holds in general for variable attrition-rate coefficients. Hence, for such cases we will have to use a different approach to develop victory-prediction conditions analogous to (4) or its special case (5). Accordingly, we observe that (4) may also be obtained from the time history of the X force level

$$x(t)=\{(x_0-y_0\sqrt{a/b})\exp(\sqrt{ab}t)+(x_0+y_0\sqrt{a/b})\exp(-\sqrt{ab}t)\}/2, \tag{6}$$

via determining the time for X to reach his breakpoint (i.e. $x(t=t_X{}^{BP})=x_{BP}$)

$$t_X{}^{BP}=(1/\sqrt{ab})\ln(\{-x_{BP}+\sqrt{x_{BP}{}^2+y_0{}^2a/b-x_0{}^2}\}/\{y_0\sqrt{a/b}-x_0\}), \tag{7}$$

---

†The influential 19th-century German military philosopher, Carl von Clausewitz (1780–1831), stated in his classic work *On War* (*Vom Kriege*) (see p. 276 of [7]), "The best Strategy is *always be to very strong*, first generally then at the decisive point. . . . There is no more imperative and no simpler law for Strategy than to *keep the forces concentrated*."

‡As pointed out in reference [26] the entire topic of modelling battle termination is a problem area on contemporary defense planning studies, and there is far from universal agreement as to even which variables should be taken as the significant variables for modeling this complex process. For further references see Taylor [23].

and requiring $t_X{}^{BP} < t_Y{}^{BP}$. The key result for obtaining (7) is that one of the two linearly independent solutions to the X force-level equation $d^2x/dt^2 - abx = 0$ is the reciprocal of the other. For a fight-to-the-finish, (7) becomes

$$t_X{}^a = \{1/(2\sqrt{ab})\} \ln (\{y_0\sqrt{a/b} + x_0\}/\{y_0\sqrt{a/b} - x_0\}), \tag{8}$$

where $t_X{}^a$ denotes the time to annihilate the X force. We observe that (5) is an immediate consequence of (8)‡.

In many applications (see Section 3 below), one is interested in whether the battle will be terminated within a given time t_g. In this case $x_0 < y_0\sqrt{a/b}$ is a necessary condition for X to be annihilated and annihilation occurs when $t_X{}^a \leq t_g$. Thus, determination of whether force annihilation will occur within a given time involves consulting a tabulation of a transcendental function, here the natural logarithm (see also (10) below). Similar results hold for other fixed force-level breakpoints.

The time history of the X force level may also be written as

$$x(t) = x_0 \cosh \sqrt{ab}\ t - y_0\sqrt{a/b} \sinh \sqrt{ab}\ t. \tag{9}$$

Taylor and Brown [25] take (9) as their point of departure for a mathematical theory for solving variable-coefficient formulations. (7) does not follow directly from (9), but (5) does via $x(t = t_X{}^a) = 0$ and

$$t_X{}^a = (1/\sqrt{ab}) \tanh^{-1}(x_0/y_0\sqrt{b/a}), \tag{10}$$

since the range of the hyperbolic tangent is [0, 1] for nonnegative arguments.

The purpose of this paper is to generalize the above to the general case of variable attrition-rate coefficients. We use (6) as our point of departure and base our development on the observation that (5) follows directly from (6), since the second term in brackets is always positive and goes to zero as $t \to +\infty$. (We will ignore the physical impossibility of negative force levels in developing results like (5)). Thus, by (6)

$$\lim_{t\to\infty} x(t) = -\infty \longleftrightarrow x_0 < \sqrt{a/b}\ y_0,$$

whence follows (5)†.

3. VARIABLE ATTRITION-RATE COEFFICIENTS.

The pioneering work of S. Bonder [4, 5] on methodology for the evaluation of military systems (in particular, mobile systems such as tanks) has generated interest in variable-coefficient

‡It is obvious that at most one of $x(t)$ and $y(t)$ can vanish. In this case we have what the mathematician calls a nonoscilliatory solution to (1) (see p. 373 of Hille [10]). Consequently, if the time to annihilate X (i.e. $t_X{}^a$ as given by (8)) is well defined, then we *must* have $y(t) > 0$ for all $t \geq 0$. Expression (5) is obtained from (8) by requiring that the argument of the natural logarithm in (8) is positive, i.e. by requiring that $t_X{}^a$ be well defined. The nonoscillation of all solutions to (11) (a special case of which is (1)) is an immediate consequence of the identity

$$x(t)y(t) = x_0y_0 - \int_0^t \{a(s)y^2(s) + b(s)x^2(s)\}ds,$$

which follows from multiplying the first equation of (11) by y, the second by x, adding, and integrating the result between 0 and t. From the above identity, it is clear that if $x(t)$ ever becomes zero, then $y(t) > 0$ for all $t \geq 0$.

Lanchester-type equations and has led to improved operations research techniques for the prediction of such coefficients (see Bonder [2, 3]; background and further references are given in Taylor and Brown [25]). Thus, we consider

$$dx/dt=-a(t)y, \qquad dy/dt=-b(t)x, \tag{11}$$

where $a(t)$ and $b(t)$ denote time-dependent Lanchester attrition-rate coefficients.

These coefficients depend on such variables as firing doctrine, firing rate, rate of target acquisition, force separation, tactical posture of targets, etc. (see [4]). Without loss of generality, we may take $a(t)=k_a g(t)$ and $b(t)=k_b h(t)$, where $g(t)$ and $h(t)$ denote time-varying factors such that $a(t)/b(t)=k_a/k_b=$constant for $g(t)=h(t)$. We will also refer to (11) as the *equations for a square-law attrition process*, since an "instaneous" square law holds even when $a(t)/b(t)$ is not constant (see Taylor and Parry [26]; also references [21, 22, and 24]).

A large class of combat situations of interest can be modeled with the following attrition-rate coefficients (see [4]).

$$a(t)=k_a(t+C)^{\mu} \qquad \text{and} \qquad b(t)=k_b(t+C+A)^{\nu}, \tag{12}$$

where A, $C\geq 0$. We will refer to these coefficients as *general power attrition-rate coefficients*. The modeling roles of A and C are discussed in Taylor and Brown [25]. We will refer to C as the *starting parameter*, since it allows us to model (with μ, $\nu\geq 0$) battles that begin within the minimum of the maximum effective ranges of the two systems. We will refer to A as the *offset parameter*, since it allows us to model (again, with μ, $\nu\geq 0$) battles between weapon systems with different ranges (i.e. opposing weapons whose fire effectiveness is "offset"). For example, let us consider Bonder's [4] constant-speed attack on a static defensive position (see also [22, 25]). Then we have

$$dx/dt=-\alpha(r)y, \qquad dy/dt=-\beta(r)x, \tag{13}$$

where $r(t)=R_0-vt$ denotes the distance (range) between the two opposing forces, R_0 denotes the battle's opening range, $v>0$ denotes the constant attack speed,

$$\alpha(r)=\begin{cases}0 & \text{for } r\geq R_\alpha,\\ \alpha_0(1-r/R_\alpha)^{\mu} & \text{for } 0\leq r\leq R_\alpha,\end{cases} \tag{14}$$

$\mu\geq 0$, and R_α denotes the maximum effective range of Y's weapon system. Similarly for $\beta(r)$, with exponent $\nu\geq 0$. In (14) the parameter μ allows us to model the range dependence of Y's fire effectiveness (see Figure 1). The offset and starting parameters are given by

$$A=(R_\beta-R_\alpha)/v, \qquad \text{and} \qquad C=(R_\alpha-R_0)/v, \tag{15}$$

and the assumption A, $C\geq 0$ implies that $R_\beta\geq R_\alpha\geq R_0$. From considering (15) and Figure 2, the reader should have no trouble understanding our terminology for A and C. In this model the de-

---

†It is clear that $x(t)\to-\infty$ implies that there exists a finite t_{X^a} such that $x(t_{X^a})=0$. By the nonoscillation of solutions to (11) (a special case of which is (1)) it follows that $y(t)>0$ for all $t\geq 0$ so that Y must win a fight-to-the-finish in finite time (see footnote ‡ p. 352).

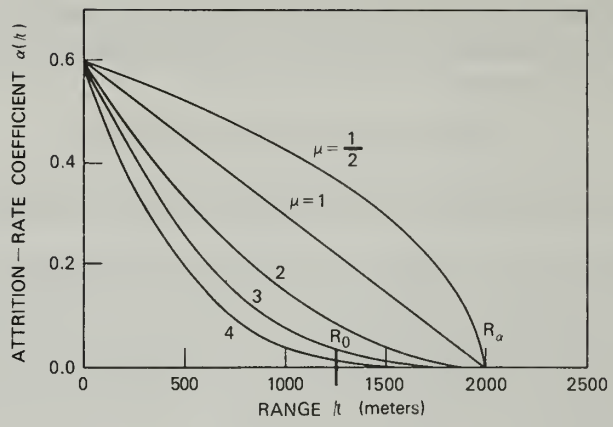


FIGURE 1. Dependence of the attrition-rate coefficient $\alpha(r)$ on the exponent μ for constant maximum effective range of the weapon system and constant kill capability at zero range. (The maximum effective range of the system is denoted $R_\alpha = 2000$ meters; $\alpha(r=0) = \alpha_0 = 0.6X$ casualties/(unit time $\times$ number of Y units) denotes the Y force weapon system kill rate at zero force separation (range). The opening range of battle is denoted as $R_0 = 1250$ meters and (as shown) $R_0 < R_\alpha$.)

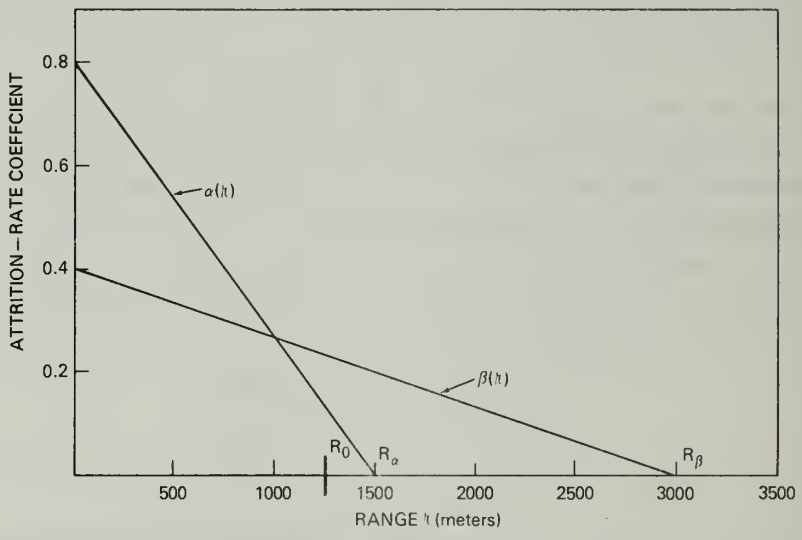


FIGURE 2. Explanation of starting parameter C and offset parameter A for power attrition-rate coefficients modelling constant-speed attack. (The maximum effective ranges of the two weapon systems are denoted as R_α and R_β. The opening range of battle is denoted as R_0 and (as shown) $R_0 <$ minimum (R_α, R_β). The starting parameter is given by $C = (R_\alpha - R_0)/v$. The offset parameter is given by $A = (R_\beta - R_\alpha)/v$.)

fensive position is overrun by the attackers (i.e. zero force separation is reached) at time $t_g = R_0/v$ and this leads to interest in predicting battle termination within a given time t_g (see Section 2 above).

Almost all previous work on the variable-coefficient equations (11) has developed infinite series solutions for force-level trajectories or represented these by tabulated functions (see [25], in particular Section 3). Relatively little attention has been given to determining the qualitative behavior of solutions to (11) (such as prediction of battle outcome) without explicitly computin

battle trajectories. Bonder and Farrell [4]†, however, have considered force annihilation within a given time. Using comparison techniques from the theory of ordinary differential equations (see, for example, Coddington and Levinson [8]), they obtained a rather strong sufficient condition for the special case of the model (13) with $\mu=\nu=1$ and $A>0$.

4. A MATHEMATICAL THEORY FOR CONDITIONS OF FORCE ANNHILATION

Motivated by the constant-coefficient results††, we introduce the *exponential-like general Lanchester functions* E_X^+, E_X^-, E_Y^+, and E_Y^-, defined by

$$\text{(16)}\qquad \begin{cases} dE_X^+/dt=\sqrt{k_b/k_a}\,a(t)E_Y^+ & \text{with } E_X^+(t=t_0)=1/Q, \\ dE_Y^+/dt=\sqrt{k_a/k_b}\,b(t)E_X^+ & \text{with } E_Y^+(t=t_0)=1, \end{cases}$$

and

$$\text{(17)}\qquad \begin{cases} dE_X^-/dt=-\sqrt{k_b/k_a}\,a(t)E_Y^- & \text{with } E_X^-(t=t_0)=1, \\ dE_Y^-/dt=-\sqrt{k_a/k_b}\,b(t)E_X^- & \text{with } E_Y^-(t=t_0)=Q, \end{cases}$$

where $Q\geq 0$ is to be determined so that E_X^- and E_Y^- have certain desired properties. We assume that $a(t)$ and $b(t)$ are defined, positive, and continuous for $t_0<t<+\infty$ with $t_0\leq 0$. Further restrictions on $a(t)$ and $b(t)$ are given by Conditions (A) and (B) below.

We want the solutions E_X^+ and E_Y^+ to (16) and E_X^- and E_Y^- to (17) to satisfy the given initial conditions, to be continuous, and to act like exponentials‡ for general $a(t)$ and $b(t)$. Accordingly, we must further restrict $a(t)$ and $b(t)$ slightly. We therefore assume that the following two conditions hold.

CONDITION (A): $\int_{t_0}^{t} a(s)ds$ and $\int_{t_0}^{t} b(s)ds$ are both bounded for all finite $t\geq t_0$.

CONDITION (B): $\lim_{t\to+\infty}\int_{t_0}^{t} a(s)ds=+\infty$, and $\lim_{t\to+\infty}\int_{t_0}^{t} b(s)ds=+\infty$.

We then have (*see* Hille [10] or Lee and Markus [15])

THEOREM 1: Condition (A) is a necessary and sufficient condition for (16) and (17) to have a continuous solution for all finite $t\geq t_0$. Also,

THEOREM 2: Condition (B) implies that both E_X^+ and E_Y^+ are unbound, e.g.

$$E_X^+(t)\to+\infty \text{ as } t\to+\infty.$$

---

†Bonder and Farrell [4] take range (i.e. force separation) to be the independent variable in their work, while Taylor [22] and Taylor and Brown [25] take time as we have done in this paper.

††Recalling the constant coefficient result (6), we consider

$$d\{\exp(\sqrt{ab}\,t)\}/dt=a\sqrt{b/a}\exp(\sqrt{ab}\,t)=b\sqrt{a/b}\exp(\sqrt{ab}\,t),$$

and

$$d\{\exp(-\sqrt{ab}\,t)\}/dt=-a\sqrt{b/a}\exp(-\sqrt{ab}\,t)=-b\sqrt{a/b}\exp(-\sqrt{ab}\,t),$$

to obtain motivation for (16) and (17).

‡We want E_X^- to behave like a decaying exponential and E_X^+ like an increasing one.

PROOF: The theorem follows from oberving that, for example, $dE_X^+/dt \geq \sqrt{k_b/k_a}\, a(t)$.

Q.E.D.

Besides satisfying (16) and (17), E_X^+ and E_X^- are linearly independent solutions to the *X force-level equation*

$$\frac{d^2x}{dt^2} - \left\{\frac{1}{a(t)}\frac{da}{dt}\right\}\frac{dx}{dt} - a(t)b(t)x = 0. \tag{18}$$

In the initial conditions of (16) and (17), $t_0 = \max(t_0^X, t_0^Y)$, where t_0^X denotes the largest finite singular point on the t-axis for (18) (*see* Taylor and Brown [25]; also p. 69 of Ince [11]).†† Since E_X^+ and E_X^- are linearly independent, we may use them to construct the general solution to (18). It follows that the solution to (11) with initial conditions (2) is given by

$$x(t) = \{[x_0 E_Y^-(t=0) - \sqrt{k_a/k_b}\, y_0 E_X^-(t=0)]E_X^+(t) + [x_0 E_Y^+(t=0) + \sqrt{k_a/k_b}\, y_0 E_X^+(t=0)]E_X^-(t)\}/2, \tag{19}$$

and

$$y(t) = \{[y_0 E_X^-(t=0) - \sqrt{k_b/k_a}\, x_0 E_Y^-(t=0)]E_Y^+(t) + [y_0 E_X^+(t=0) + \sqrt{k_b/k_a}\, x_0 E_Y^+(t=0)]E_Y^-(t)\}/2,$$

where we have made use of the easily verifiable fact that (*see* [25])

$$E_X^+(t)E_Y^-(t) + E_X^-(t)E_Y^+(t) = 2 \qquad \forall t. \tag{20}$$

Let us now consider how to choose the general Lanchester functions E_X^- and E_X^+ so that they play the roles of a decaying exponential and an increasing one. Force-annihilation-prediction conditions may then be obtained from (19) by inspection (again, *see* footnote, p. 353). We recall the constant-coefficient result (6) and its consequence (5), obtained using

$$\lim_{t \to +\infty} \exp(-\sqrt{ab}\, t) = 0 \text{ and } \exp(\pm\sqrt{ab}\, t) > 0.$$

Clearly, $E_X^+(t) > 0$ for all $t \geq t_0$ when $Q \geq 0$. We have already shown that Condition (B) implies that $E_X^+(t)$, which satisfies (16), grows without bound just as an increasing exponential does. We will now show how to choose E_X^- so that it corresponds to a decaying exponential. Similar statements hold for E_Y^- and E_Y^+. Considering (17), we see that we should choose E_X^- and E_Y^- to remain positive for all t so that by (17) they continuously decrease. Furthermore, we will be able to specify such behavior for E_X^- and E_Y^- by our selection of the parameter Q in the initial conditions for (17).

The solution $E_X^-(t)$, $E_Y^-(t)$ to (17) depends continuously on the parameter Q of the initial conditions (*see*, for example, Hille [10]). We denote this dependence by $E_X^-(t; Q)$, $E_Y^-(t; Q)$. Let $Q^* = Q^*(a(t), b(t))$ denote the unique (see [10]) value of Q such that

$$E_X^-(t; Q=Q^*),\ E_Y^-(t; Q=Q^*) > 0 \text{ for all finite } t \geq t_0. \tag{21}$$

---

††We take $a(t)$ and $b(t)$ to be analytic in the (finite) complex plane except for a finite number of singularities on the real axis. The singularities of (18) then occur at the zeros and singularities of $a(t)$ and at the singularities of $a(t)b(t)$. Consequently, t_0 belongs to the set of points consisting of the zeros and singularities of $a(t)$ and $b(t)$ (see Taylor and Brown [25]). We define t_0 this way in order to reduce the number of tabulations of exponential general Lanchester functions required for force-annihilation analyses (see Theorem 4). For example, for the general power attrition-rate coefficients (12) we have $t_0 = -C$, and for fixed $A \geq 0$, μ, and ν only a single tabulation of e_X^- and e_Y^- is required to handle all problems with $C \geq 0$ (see, for example, Theorem 5).

Using arguments similar to those given by Hille (see pp. 437–439 of [10]), one can show that†

$$\lim_{t\to+\infty} E_X^-(t;\, Q^*)=\lim_{t\to+\infty} E_Y^-(t;\, Q^*)=0. \tag{22}$$

It is intuitively obvious that such a Q^* exists, and we will prove its existence in some particular cases below. As we shall see, knowledge of Q^* provides valuable information about the qualitative behavior of force-level trajectories for the Lanchester-type equations (11). Let us refer to the problem of determining Q^* such that (22) holds as the *auxiliary parity-condition problem*. Unless explicitly stated otherwise, for convenience we will denote, for example, $E_X^+(t;\, Q^*)$ as $E_X^+(t)$.

COMMENT 1: For a constant ratio of attrition-rate coefficients, i.e.

$$a(t)=k_a h(t), \qquad \text{and} \qquad b(t)=k_b h(t), \tag{23}$$

where $h(t)$ denotes the common time-varying factor of the two coefficients, it readily follows from the results given in references [4] and [20] that $Q^*=1$ and

$$E_X^+(t)=E_Y^+(t)=\exp\{\psi(t)\}, \text{ and } E_X^-(t)=E_Y^-(t)=\exp\{-\psi(t)\}, \tag{24}$$

where

$$\psi(t)=\sqrt{k_a k_b}\int_{t_0}^{t} h(s)ds.$$

COMMENT 2: By Theorem 1 of Taylor and Brown [25], the solution (19) simplifies to the form of (6) only if (23) holds.

The above exponential-like general Lanchester functions may be related to Taylor and Brown's [25] hyperbolic-like general Lanchester functions. Let

$$C_X(t)=x_1(t),\ S_X(t)=x_2(t),\ C_Y(t)=y_1(t), \text{ and } S_Y(t)=y_2(t), \tag{25}$$

where x_1, x_2, y_1, and y_2 denote the general Lanchester functions introduced by Taylor and Brown. Then, similar to the well-known relationships between the hyperbolic and exponential functions, we have

$$\begin{aligned} C_X(t)&=\{Q^*E_X^+(t)+E_X^-(t)\}/2, & S_X(t)&=\{Q^*E_X^+(t)-E_X^-(t)\}/(2Q^*),\\ C_Y(t)&=\{Q^*E_Y^+(t)+E_Y^-(t)\}/(2Q^*), & S_Y(t)&=\{Q^*E_Y^+(t)-E_Y^-(t)\}/2. \end{aligned} \tag{26}$$

The determination of Q* will be slightly simplified for general power attrition-rate coefficients (12) by considering a modified auxiliary parity-condition problem. For this purpose we introduce the new independent variable

$$s=K\sqrt{k_b/k_a}\int_{t_0}^{t} a(\sigma)d\sigma, \tag{27}$$

and define $s_0=s(t=0)\geq 0$ for $t_0\leq 0$. K is an, at present, undetermined parameter. It will be chosen so that a more convenient canonical system of differential equations arises in the modified auxiliary

†From (17) it is clear that E_X^- and E_Y^- are continuously decreasing when (21) holds.

parity-condition problem. By Condition (A), the transformation (27) is well defined for $t \geq t_0$. It has an inverse $t(s)$, since $a(t) > 0 \forall t > t_0$. Letting

(28) $e_X^+(s) = KE_X^+(t(s))$, $e_Y^+(s) = E_Y^+(t(s))$, $e_X^-(s) = E_X^-(t(s))$, and $e_Y^-(s) = E_Y^-(t(s))/K$,

the substitution (27) transforms (16) through (18) into

(29) $$\begin{cases} de_X^+/ds = e_Y^+ & \text{with } e_X^+(s=0) = 1/Z, \\ de_Y^+/ds = I(s)e_X^+ & \text{with } e_Y^+(s=0) = 1, \end{cases}$$

(30) $$\begin{cases} de_X^-/ds = -e_Y^- & \text{with } e_X^-(s=0) = 1, \\ de_Y^-/ds = -I(s)e_X^- & \text{with } e_Y^-(s=0) = Z, \end{cases}$$

and

(31) $$d^2x/ds^2 - I(s)x = 0,$$

where for any Q

(32) $$Z = Q/K,$$

and

(33) $$I(s) = (\{b(t)/k_b\}/\{a(t)/k_a\})/K^2,$$

is the invariant of the normal form (31) (see p. 119 of Kamke [12]) and $t = t(s)$ by (27). The parameter K will be chosen to simplify the form of $I(s)$. In our later work the equation (31) will be easier to analyze than (18).

We will refer to the problem of determining $Z^* = Z^*(a(t), b(t))$ such that

(34) $$e_X^-(s; Z = Z^*),\ e_Y^-(s; Z = Z^*) > 0 \text{ for all finite } s \geq 0,$$

as the modified auxiliary parity-condition problem. By (32) we then have that $Q^* = KZ^*$.

We also observe that the solution to (31) which satisfies (2) may be expressed in terms of these exponential-like general Lanchester functions for $s \geq s_0$ as

(35) $$x(s) = \{[x_0 K e_Y^-(s=s_0) - \sqrt{k_a/k_b}\, y_0 e_X^-(s=s_0)]e_X^+(s)/K + [x_0 K e_Y^+(s=s_0) + \sqrt{k_a/k_b}\, y_0 e_X^+(s=s_0)]e_X^-(s)/K\}/2,$$

and similarly

(36) $$y(s) = \{[y_0 e_X^-(s=s_0) - \sqrt{k_b/k_a}\, x_0 K e_Y^-(s=s_0)]e_Y^+(s) + [y_0 e_X^+(s=s_0) + \sqrt{k_b/k_a}\, x_0 K e_Y^+(s=s_0)]e_Y^-(s)\}/2,$$

where by (20) and (28)

(37) $$e_X^+(s)e_Y^-(s) + e_X^-(s)e_Y^+(s) = 2 \qquad \forall s.$$

From our choice of Q^* such that (21) and (22) hold, we can immediately infer the behavior of the solution (19) to (11) as $t \to +\infty$ and similarly for $y(t)$. Thus, we have

THEOREM 3: $\lim_{t \to +\infty} x(t) = -\infty$ if and only if

$$x_0/y_0 < \sqrt{k_a/k_b}\, E_X^-(t=0; Q^*)/E_Y^-(t=0; Q^*).$$

Equivalently, we may state (see footnote, p. 353 again)

THEOREM 4: Consider combat between two homogeneous forces described by (11). Assume that (11) applies for all time and that Y "wins" when $x(t_f)=0$ with $y(t_f)>0$. Then, Y will win if and only if

$$x_0/y_0<\sqrt{k_a/k_b}\,E_X^{-}(t=0;\,Q^*)/E_Y^{-}(t=0;\,Q^*).$$

Thus, we see that tabulations of such new exponential-like general Lanchester functions $E_X^{-}(t;\,Q^*)$ and $E_Y^{-}(t;\,Q^*)$ would facilitate force-annihilation prediction. By our choice of t_0 only one such tabulation is necessary for given attrition-rate coefficients $a(t)$ and $b(t)$ (see Note 9). Alternatively, we may express the force-annihilation condition of Theorem 4 in terms of Taylor and Brown's [25] hyperbolic-like general Lanchester functions (see equations (25) and (26)). We have then

THEOREM 4′: Consider combat between two homogeneous forces described by (11). Assume that (11) applies for all time and that Y "wins" when $x(t_f)=0$ with $y(t_f)>0$. Then, Y will win if and only if

$$x_0/y_0<\sqrt{k_a/k_b}\,\{C_X(t=0)-Q^*S_X(t=0)\}/\{Q^*C_Y(t=0)-S_Y(t=0)\}.$$

As an immediate corollary to Theorem 4 we have

COROLLARY 4.1: For $t_0=0$, Y will win if and only if

$$\frac{x_0}{y_0}<\frac{1}{Q^*}\sqrt{\frac{k_a}{k_b}}.$$

One shortcoming of our above development is that Theorem 4 and its corollaries are basically existence theorems for the annihilation of one side by the other at some unknown future point in time. As the enemy initial force level decreases towards parity (i.e. equality holding in (38)), the time required to annihilate a force becomes larger and larger. There is, in fact, no limit to how large it may become. Moreover, there is a large class of tactically significant battles (see Bonder's [4] constant-speed attack on a static defensive position discussed in Section 3 above) which has a built-in time limit, denoted as t_g. Hence, it would be desirable to have a method for determining from initial conditions (without explicitly computing the entire force-level trajectories) whether or not force annihilation will occur within a given finite time t_g. For our general model (11), Theorem 4 tells us that

$$x_0/y_0<\sqrt{k_a/k_b}\;E_X^{-}(t=0;\,Q^*)/E_Y^{-}(t=0;\,Q^*), \tag{38}$$

is a necessary condition for $x(t_X{}^a)=0$ with $t_X{}^a\leq t_g$, where $t_X{}^a$ denotes the time at which the X force is annihilated. Motivated by the well-known constant-coefficient result (8), one intuitively sees that determining whether or not $t_X{}^a\leq t_g$ will require the appropriate tabulations of new transcendents (i.e. new functions). Different such functions arise from different fundamental systems chosen to construct the solution to, for example, the X force-level equation (18). Let us now investigate which fundamental system of solutions is the most useful.

We begin our investigation by outlining for the new exponential-like general Lanchester functions how to determine from initial conditions (without explicitly computing the force-level trajectories) whether or not force annihilation will occur within a given finite time t_g. Looking at (19)

and setting $x(t)=0$, we see that we must solve for $E_X^+(t)/E_X^-(t)=\eta(t)$†. For any other fundamental system (i.e. pair) of solutions, we must still solve for such a quotient (cf. the constant-coefficient results (8) and (10)). Thus, our force-annihilation determination requires the use of tabulations of the quotient of two linearly independent solutions to, for example, the X force-level equation. Our question is now which quotient is the most useful (i.e. which fundamental system yields the most useful quotient). We will show that the quotient of two exponential-like general Lanchester functions is not numerically satisfactory for such determinations, whereas the quotient of two hyperbolic-like general Lanchester functions is.

Motivated by the result for a constant ratio of attrition-rate coefficients (23) that

$$\eta(t)=\exp\{2\psi(t)\},$$

we use the following notation for the $\eta(t)$ of the X force-level equation

$$E_{2X}^+(t)=E_X^+(t)/E_X^-(t), \tag{39}$$

with $E_{2Y}^+(t)$ being similarly defined. Assuming that (38) holds, we see from (19) that if $x(t=t_X^a)=0$, then

$$E_{2X}^+(t_X^a)=\{\sqrt{k_a/k_b}\,y_0E_X^+(t=0;\,Q^*)+x_0E_Y^+(t=0;\,Q^*)\}/\{\sqrt{k_a/k_b}\,y_0E_X^-(t=0;\,Q^*)-x_0E_Y^-(t=0;\,Q^*)\}.$$

We will prove below that $E_{2X}^+(t)=E_{2X}^+(t;\,Q^*)$ is a strictly increasing function of t with initial value $1/Q^*$ at $t=t_0$. Consequently, the inverse function $E_{2X}^{+-1}(\xi)$ is well defined $\forall\xi\epsilon[1/Q^*,+\infty)$, and in this case the time to annihilate X is given by

$$t_X^a=E_{2X}^{+-1}(\{\sqrt{k_a/k_b}\,y_0E_X^+(t=0;\,Q^*)+x_0E_Y^+(t=0;\,Q^*)\}/\{\sqrt{k_a/k_b}\,y_0E_X^-(t=0;\,Q^*)-x_0E_Y^-(t=0;\,Q^*)\}). \tag{40}$$

However, in general for $x_{BP}>0$ we have been unable to develop an analogous result‡.

We now show that $E_{2X}^+(t)$ (defined by (39)) is a strictly increasing function. We readily compute using (16), (17), and (20), that

$$dE_{2X}^+/dt=2\sqrt{k_b/k_a}\,a(t)/\{E_X^-(t)\}^2 \qquad \text{with } E_{2X}^+(t=t_0)=1/Q^*, \tag{41}$$

whence follows the monotonicity. Similar results hold for the modified exponential-like general Lanchester functions defined by (29) and (30).

---

†It is well known (see, for example, pp. 647–650 of Hille [10] or p. 120 of Kamke [12]) that the quotient of two linearly independent solutions to (31), which is equivalent to (18), satisfies Schwarz's (third order) differential equation (see Schwarz [18])

$$\{\eta,\,s\}=-2I(s),$$

where η denotes the quotient of two linearly independent solutions to (31) [i.e. $\eta=e_X^+(s)/e_X^-(s)$],

$$\{\eta,\,s\}=\eta'''-(3/2)(\eta'')^2/\eta'$$

denotes the Schwarzian derivative of η with respect to s, η' denotes $d\eta/ds$, etc., and $I(s)$ denotes the invariant of the normal form (31). For numerical computation of η, however, there is no advantage to consider this third order equation, and it is preferable to calculate η from, for example, $\eta(s)=e_X^+(s)/e_X^-(s)$, where e_X^+ and e_X^- also satisfy (29) and (30).

‡Recalling our development of (7), we see that, except for the special case in which (23) holds, this same approach fails to yield the time for X to reach his breakpoint (assumed to be positive)

$$(\text{i.e. } t_X^{BP} \text{ such that } x(t=t_X^{BP})=x_{BP}>0).$$

Consequently, it is apparently impossible to predict in the manner described in the main text the outcome of a fixed force-level breakpoint battle with positive breakpoints unless (23) holds.

Thus, (I) determination of $Q^*=Q^*(a(t), b(t))$, and (II) tabulation of $E_{2X}{}^+(t)$ and $E_{2Y}{}^+(t)$ would allow one to determine whether or not force annihilation occurs in such battles with finite time limit without explicitly computing the entire force-level trajectories (see Figures 3, 4, and 5 of [25]). Unfortunately, there is a serious drawback to considering $E_{2X}{}^+(t)$ and $E_{2Y}{}^+(t)$: accurate tabulations are difficult to generate (in fact, they are essentially impossible for large values of t), since both functions are basically increasing exponentials so that any error in their initial value $1/Q^*$ (which, in general, can be only approximately numerically determined) becomes tremendously magnified over time.

We may develop numerically satisfactory functions for prediction of force annihilation within a given finite time, however, by considering the hyperbolic-like general Lanchester functions of Taylor and Brown [25]. Let us therefore define

$$T_X(t)=S_X(t)/C_X(t). \tag{42}$$

The functions $T_X(t)$ and $T_Y(t)$ are analogous to the hyperbolic tangent, to which they reduce for a constant ratio of attrition-rate coefficients. Considering (25) and equation (16) of [25], we see that, for example, $T_X(t)$ does not depend on Q^*, since $S_X(t)$ and $C_X(t)$ do not. Thus, $T_X(t)$ and $T_Y(t)$ are numerically suitable for determining whether or not force annihilation will occur within a given finite time.

Using the results given in Table I of Taylor and Brown [25], we readily compute that

$$dT_X/dt=\sqrt{k_b/k_a}\,a(t)/\{C_X(t)\}^2 \qquad \text{with } T_X(t=t_0)=0. \tag{43}$$

Hence, $T_X(t)$ is a strictly increasing function, and its inverse $T_X{}^{-1}$ is well defined. Let us now establish an upper bound for $T_X(t)$. By (26) and (42), we have

$$T_X(t)=(1/Q^*)\{Q^*E_X{}^+(t)-E_X{}^-(t)\}/\{Q^*E_X{}^+(t)+E_X{}^-(t)\},$$

whence it follows that for $t\geq t_0$

$$0\leq T_X(t)<1/Q^*, \qquad \text{with } \lim_{t\to+\infty} T_X(t)=1/Q^*. \tag{44}$$

Thus, our current investigation has yielded important information about the asymptotic behavior of hyperbolic-like general Lanchester functions.

To determine $t_X{}^a$ such that $x(t_X{}^a)=0$, we write the solution to (18) which satisfies the initial conditions (2) as [25]

$$x(t)=x_0\{C_Y(t=0)C_X(t)-S_Y(t=0)S_X(t)\}-y_0\sqrt{k_a/k_b}\{C_X(t=0)S_X(t)-S_X(t=0)C_X(t)\}$$

and find that when (38) holds, $t_X{}^a$ is given by

$$t_X{}^a=T_X{}^{-1}(\{x_0C_Y(t=0)+y_0\sqrt{k_a/k_b}S_X(t=0)\}/\{y_0\sqrt{k_a/k_b}C_X(t=0)+x_0S_Y(t=0)\}). \tag{45}$$

We observe that by (26) the argument of the inverse function $T_X{}^{-1}$ in (45) belongs to the range of T_X (see (44)) when (38) holds (see Theorem 4′). For $t_0=0$, (45) simplifies to

$$t_X{}^a=T_X{}^{-1}(x_0/y_0\sqrt{k_b/k_a}).$$

Whether or not force annihilation occurs within a given finite time t_g then depends on whether or not $t_X{}^a \leq t_g$. Thus, (I) determination of Q^*, and (II) tabulation of $T_X(t)$ and $T_Y(t)$ would allow one to determine (without explicitly computing the force-level trajectories) the time at which a side is annihilated.

5. APPLICATION TO POWER ATTRITION-RATE COEFFICIENTS

Let us now apply the above general theory to (11) with the general power attrition-rate coefficients (12). We observe that in this case $t_0 = -C$, where $C \geq 0$. In order that Condition (A) holds we must have $\mu, \nu > -1$, and then Condition (B) is satisfied.

As we have seen in Section 4, our theory of force-annihilation prediction depends on knowing Q^*, the solution to the auxiliary parity-condition problem. For the power attrition-rate coefficients (12), it is more convenient, however, to determine Q^* via the modified auxiliary parity-condition problem (30) (see also (34)). Hence, we apply the transformations (27) and (28) to (17). For the coefficients (12), equations (30) take the form

$$(46) \qquad \begin{cases} de_X{}^-/ds = -e_Y{}^- & \text{with } e_X{}^-(s=0)=1, \\ de_Y{}^-/ds = -s^\beta(1+\gamma/s^\alpha)^\nu e_X{}^- & \text{with } e_Y{}^-(s=0)=Z, \end{cases}$$

where the parameter K in (27) is given by $K=(\sqrt{k_a k_b}/(\mu+1))^{2p-1}$; and we have $p=(\mu+1)/(\mu+\nu+2)$, $q=1-p$, $\alpha=1/(\mu+1)$, $\beta=(\nu-\mu)/(\mu+1)$, and $\gamma=A(\sqrt{k_a k_b}/(\mu+1))^{2/(\mu+\nu+2)}$. For $\mu, \nu > -1$, we have $0<p, q<1$.

After we have solved the above modified auxiliary parity-condition problem (i.e. determined $Z=Z^*$ for (46) such that (34) holds), we have all the information required to determine, without explicitly computing the entire force-level trajectories, whether or not force annihilation occurs in battles modelled with (12). We may apply Theorem 4 via (28) (possibly using (26)) to see who can be annihilated and use results such as (40) (or equivalent ones for hyperbolic-like power Lanchester functions (see Taylor and Brown [25])) to see if force annihilation occurs in a given finite time (for example, for battles modelled by (13)). When C=0 for the coefficients (12) (e.g. for the model (13), $R_0 = R_\alpha \leq R_\beta$ from (15)), we have by Corollary 4.1

COROLLARY 4.2: For combat between two homogeneous forces modelled by (11) and (12) with $C=0$, the X force can be annihilated if and only if

$$x_0/y_0 < (1/Z^*)(\sqrt{k_a k_b}/(\mu+1))^{1-2p}\sqrt{k_a/k_b},$$

where $Z^* = Z^*(\gamma, \mu, \nu)$ is such that (34) holds for (46).

We will next give exact analytic results for cases of no offset (i.e. $A=0 \Rightarrow \gamma=0$) and discuss the difficulties of determining Z^* when there is offset (i.e. A, $\gamma>0$). Moreover, results for the special case of no offset help provide a lower bound for Z^* in the general case.

6. RESULTS FOR POWER ATTRITION-RATE COEFFICIENTS WITH NO OFFSET

When the offset parameter $A=0$, equations (46) become for $\mu, \nu > -1$

$$\text{(47)} \qquad \begin{cases} de_X^-/ds = -e_Y^- & \text{with } e_X^-(s=0)=1, \\ de_Y^-/ds = -s^\beta e_X^- & \text{with } e_Y^-(s=0)=Z. \end{cases}$$

Solving (47) by successive approximations (see pp. 757–760 of Taylor [22]), we obtain

$$\text{(48)} \qquad e_X^-(s;Z)=\sum_{k=1}^{\infty} p^{2k}s^{k(\beta+2)} \Big/ \left\{ \prod_{j=0}^{k} j(j-p) \right\} - Z \sum_{k=0}^{\infty} p^{2k}s^{k(\beta+2)+1} \Big/ \left\{ \prod_{j=1}^{k} j(j+p) \right\},$$

which is not the most useful result for large s. Without explicitly computing $e_X^-(s;Z)$ it is impossible to determine from (48) how $e_X^-(s;Z)$ behaves for increasing s. However, let us write $e_X^-(s;Z)$ as

$$\text{(49)} \qquad e_X^-(s;Z)=\Gamma(q)p^{-q}\{A_\beta(s)+[1-Zp^{q-p}\Gamma(p)/\Gamma(q)]ps^{1/2}I_p(S)\},$$

(observe that e_X^- satisfies the generalized Airy equation (see Swanson and Headley [19])

$$d^2e_X^-/ds^2 - s^\beta e_X^- = 0,$$

which is well known to be reducible to Bessel's equation) for any arbitrary Z, where $A_\beta(s)$ denotes the generalized Airy function of the first kind of order β (see Swanson and Headley [19]), $I_p(S)$ denotes the modified Bessel function of the first kind of order p, and $S=2ps^{(\beta+2)/2}$. Also,

$$\text{(50)} \qquad e_Y^-(s;Z)=\Gamma(q)p^{-q}\{(2p/\pi)\ (\sin p\pi)s^{(\beta+1)/2}K_q(S)-[1-Z_p{}^{q-p}\Gamma(p)/\Gamma(q)]dF/ds(s)\},$$

where $K_q(S)$ denotes the modified Bessel function of the third kind (also called Macdonald's function) of order q and $F(s)=ps^{1/2}I_p(S)$.

The behavior of $A_\beta(s)$ for $s \geq 0$ is readily seen from (see Swanson and Headley [19])

$$A_\beta(s)=(2p/\pi)(\sin p\pi)s^{1/2}K_p(S)$$

It is readily seen (see pp. 119 and 123 of Lebedev [14] or pp. 250–251 of Olver [17]) that $K_p(S)$ is strictly decreasing and positive, and

$$\lim_{s\to+\infty} s^\nu K_p(S)=0,$$

where ν is any real number. It is then clear (see also p. 1404 of [19] that

$$A_\beta(s),\ I_p(S)>0\ \forall s\geq 0, \qquad \lim_{s\to+\infty} A_\beta(s)=0, \qquad \text{and } \lim_{s\to+\infty} I_p(S)=+\infty.$$

Consequently, the requirement that (34) holds (i.e. that e_X^- and e_Y^- behave like strictly decaying exponentials) means that the second term in the expressions (49) and (50) must vanish. Thus, for $\gamma=0$ we have

$$\text{(51)} \qquad Z^*(\gamma=0,\mu,\nu)=p^{p-q}\Gamma(q)/\Gamma(p).$$

Hence, Theorem 4 becomes †:

THEOREM 5: Consider combat between two homogeneous forces described by (11) with attrition-rate coefficients (12) with $A=0$. Assume that the model applies for all time and that Y "wins" when $x(t_f)=0$ with $y(t_f)>0$. Then Y will win if and only if

$$x_0/y_0<\sqrt{k_a/k_b}(\sqrt{k_ak_b}/(\mu+1))^{q-p}e_X^-(s=s_0;\ Z^*)/e_Y^-(s=s_0;\ Z^*),$$

where

$$Z^*=Z^*(\gamma=0,\ \mu,\ \nu)=p^{p-q}\Gamma(q)/\Gamma(p),$$

$$e_X^-(s;\ Z^*)=p^{-p}\Gamma(q)A_\beta(s),$$

and

$$e_Y^-(s;\ Z^*)=(2p/\pi)(\sin p\pi)p^{-q}\Gamma(q)s^{(\beta+1)/2}K_q(S).$$

For $C=0$, we have $s_0=0$ so that Y will win if and only if

$$x_0/y_0<\sqrt{k_a/k_b}(\sqrt{k_ak_b}/(\mu+\nu+2))^{q-p}\Gamma(p)/\Gamma(q).$$

We observe that the infinite series form (48) is not of any value for determining asymptotic properties of the solution e_X^- to (47) (and consequently Z^*), although it is useful for computational purposes.

7. OFFSET LINEAR ATTRITION-RATE COEFFICIENTS

When the offset parameter $A>0$, explicit analytic results for Z^* are apparently not possible. Before considering the general case of $\mu,\ \nu>-1$ and $\gamma>0$, we will find it instructive to consider the special case of offset linear attrition-rate coefficients (i.e. $\gamma>0$ with $\mu=\nu=1$) studied by Bonder and Farrell [4]. This will show us why analytic results for Z^* are elusive in cases of positive offset.

For $\gamma>0$ with $\mu=\nu=1$, equations (30) and (31) become

$$d^2x/ds^2-(1+\gamma/\sqrt{s})x=0, \tag{52}$$

---

† For the case of power attrition-rate coefficients with no offset (i.e. $A=0$ in (12)), the second annihilation condition given in Theorem 5 (i.e. the one for $C=0$) and an equivalent form of the first (i.e. the one for $C\geq 0$) may be developed by inspection when one expresses, for example, the time history of the X force level (which satisfies (18)) in terms of the so-called generalized Airy functions (see Swanson and Headley [19]). For example, for $C=0$ we have

$$x(t)=(p^p/(2p))\{X+Y\}A_\beta(T)+(p^p/(2\sqrt{p}))\{X-Y\}B_\beta(T),$$

where A_β and B_β denote the generalized Airy functions of the first and second kinds of order β,

$$X=x_0\Gamma(q),\qquad Y=y_0\sqrt{k_ak_b}\,\Gamma(p)(\sqrt{k_ak_b}/(\mu+\nu+2))^{1-2p},$$

and

$$T=(\sqrt{k_ak_b}/(\mu+1))^{2p}t^{\mu+1}.$$

The result given in Theorem 5 for $C=0$ follows from the properties of the generalized Airy functions

$$(\text{i.e. } A_\nu(\xi),\ B_\nu(\xi)>0\ \forall\ \xi\geq 0,\qquad \lim_{\xi\to+\infty}A_\nu(\xi)=0,\qquad \text{and } \lim_{\xi\to+\infty}B_\nu(\xi)=+\infty)$$

and the above representation for $x(t)$. Unfortunately, this result does not generalize to other cases of interest, although it did motivate our general theory of force annihilation developed in this paper. The generalized Airy functions may be considered to be generalizations of the exponential function (see p. 446 of [1] or p. 393 of [17] for plots of the standard (i.e. $\beta=1$) Airy functions) and arise in the study of the asymptotic behavior of solutions to certain differential equations.

with initial conditions $x(s=0)=x_0$ and $dx/ds(s=0)=-\sqrt{k_a/k_b}\,y_0$, and

$$(53)\qquad \begin{cases} de_X^-/ds=-e_Y^- & \text{with } e_X^-(s=0)=1, \\ de_Y^-/ds=-(1+\gamma/\sqrt{s})e_X^- & \text{with } e_Y^-(s=0)=Z. \end{cases}$$

Let us now consider solving the above modified auxiliary parity-condition problem (i.e. determining $Z=Z^*$ for (53) such that (34) holds). Unfortunately, (52) does not correspond to any set of special functions we can find (see footnote‡ p. 350). However, solving (53) by successive approximations, we may write

$$(54)\qquad e_X^-(s)=h(s)-Zw(s), \text{ and } e_Y^-(s)=ZH(s)-W(s),$$

where

$$h(s)=h(s;\gamma)=h(\lambda=s,\ \mu=\gamma/\sqrt{s}),$$

$w(s)$, $H(s)$, and $W(s)$ denote the auxiliary offset linear Lanchester functions introduced by Taylor and Brown [25]. Infinite series representations of these functions are given in [25]. Subsequent research has shown that these hyperbolic-like Lanchester functions possess the following properties:

PROPERTY 1: $dh/ds=W$, $\quad dW/ds=(1+\gamma/\sqrt{s})h$

PROPERTY 2: $dw/ds=H$, $\quad dH/ds=(1+\gamma/\sqrt{s})w$

PROPERTY 3: $h(s)H(s)-w(s)W(s)=1 \quad \forall s$

PROPERTY 4: $h(s=0)=H(s=0)=1$

PROPERTY 5: $w(s=0)=W(s=0)=0$

PROPERTY 6: $h(s;\gamma=0)=H(s;\gamma=0)=\cosh s$

PROPERTY 7: $w(s;\gamma=0)=\sinh s=W(s;\gamma=0)$.

Unfortunately, information on the asymptotic behavior of the auxiliary offset linear Lanchester functions for large $s>0$, which is needed to solve the modified auxiliary parity-condition problem (53), is lacking and apparently not obtainable by the standard methods involving integral representation (see Ince [11] or Oliver [17]). Consequently, we have not been able to develop an explicit analytic expression for $Z^*(\gamma>0,\ \mu=1,\ \nu=1)$, although we give upper and lower bounds for Z^* in the next section for general $\mu,\ \nu>-1$. Additionally, we should point out that there are computational difficulties in searching for Z^* via its definition (34): (a) one doesn't know how large to take s for "satisfactory" results, and (b) numerical difficulties in evaluating $e_X^-(s;Z)$ and $e_Y^-(s;Z)$ as given by (54) occur for large values of s (since we are taking the difference of two very large numbers and, at least on a digital computer, can retain only a limited number of significant digits in these numbers).

Equation (52) looks deceptively simple. Using variation of parameters, we may also express its solution as

$$(55)\qquad x(s)=x_0\cosh s-y_0\sqrt{k_a/k_b}\sinh s+\gamma\int_0^s \sinh(s-\sigma)x(\sigma)/\sqrt{\sigma}\,d\sigma.$$

Although (55) is a simple looking expression, this Volterra integral equation is, unfortunately, no easier to solve than (52) and leads to the same results as given by Taylor [22] and Taylor and Brown [25].

8. BOUND ON Z^* FOR POWER ATTRITION-RATE COEFFICIENTS WITH POSITIVE OFFSET

We will now develop upper and lower bounds for $Z^*(\gamma>0, \mu, \nu)$. These bounds establish the existence of Z^* (and consequently Q^*) for general power attrition-rate coefficients (12) by the continuous dependence of solutions to (46) on the initial conditions.

The following two lemmas will be used to obtain an upper bound for $Z^*(\gamma, \mu, \nu)$ for $\mu, \nu>-1$.

LEMMA 1: For $\delta\geq 1$ and $x, y>0$, $2^{\delta-1}\{x^\delta+y^\delta\}\geq(x+y)^\delta$.

PROOF: For $\delta\geq 1$, $f(x)=x^\delta$ is a convex function. A well-known theorem for convex functions says that

$$\{f(x)+f(y)\}/2\geq f([x+y]/2),$$

whence follows the lemma. Q.E.D.

LEMMA 2: For $\delta\leq 1$ and $x, y>0$, $x^\delta+y^\delta\geq(x+y)^\delta$.

PROOF: Dividing by $(x+y)^\delta$, we need to show that $[x/(x+y)]^\delta+[y/(x+y)]^\delta\geq 1$. If x and $y>0$, then $x/(x+y)$ and $y/(x+y)$ are <1. Hence, for any $\delta\leq 1$ we have $[x/(x+y)]^\delta\geq x/(x+y)$ so that

$$[x/(x+y)]^\delta+[y/(x+y)]^\delta\geq[x/(x+y)]+[y/(x+y)]=1. \quad \text{Q.E.D.}$$

Using the above lemmas, we now prove Theorem 6. The upper bound to be given in Theorem 6 might be improved upon, although we feel that for computer determination of Z^* by interval search it is not essential to have a better bound.

THEOREM 6: For $\mu>-1$ and $\nu\geq 1$, we have

$$Z^*(\gamma, \mu, \nu)<1+2^{\nu-1}(\mu+1)^2/[(\nu+1)(\mu+\nu+2)]+\gamma^\nu 2^{\nu-1}(\mu+1)^2/(\mu+2);$$

while for $\mu>-1$ and $-1<\nu\leq 1$, we have

$$Z^*(\gamma, \mu, \nu)<1+(\mu+1)^2/[(\nu+1)(\mu+\nu+2)]+\gamma^\nu(\mu+1)^2/(\mu+2).$$

PROOF: Recalling (34), we have from the first equation of (46) that $e_X^-(s; Z^*)<1$ for $s>0$; and from the second equation of (46) we then obtain for $s>0$ and $\nu\geq 1$

$$de_Y^-/ds>-s^\beta(1+\gamma/s^\alpha)^\nu\geq-2^{\nu-1}s^\beta(1+\gamma^\nu/s^{\alpha\nu}), \tag{56}$$

the latter inequality being a consequence of Lemma 1. From (30) we obtain

$$e_Y^-(s)>Z^*-2^{\nu-1}s^{(\nu+1)/(\mu+1)}\cdot(\mu+1)/(\nu+1)-2^{\nu-1}\gamma^\nu(\mu+1)s^{1/(\mu+1)}. \tag{57}$$

Using (57) and considering the first equation of (46), we obtain $e_X^-(s; Z^*)<U(s; Z^*)$, where

$$U(s;Z^*)=1+2^{\nu-1}\{(\mu+1)^2/(\mu+2)\}\{s^{(\mu+\nu+2)/(\mu+1)}\cdot(\mu+2)/[(\nu+1)(\mu+\nu+2)]+\gamma^\nu s^{(\mu+2)/(\mu+1)}\}-Z^*s.$$

Since we must have $0<e_X^-(s; Z^*)<U(s; Z^*)$ for all $s\geq 0$, it follows that for $s=1$ we must have $U(s=1; Z^*)>0$, whence follows the theorem for $\nu\geq 1$. Lemma 2 and similar arguments are used to prove the theorem for $-1<\nu\leq 1$. Q.E.D.

Let us now consider the development of a lower bound for $Z^*(\gamma>0, \mu, \nu)$. Before proving the key lemma (Lemma 3) for the proof of Theorem 7, we discuss some preliminary considerations.

As shown by Taylor and Parry [26], force annihilation for square-law attrition processes sometimes may be predicted by considering the force-ratio equation. For equations (46), the "force ratio" $u=e_X^-/e_Y^-$ satifies the Riccati equation

$$du/ds=s^\beta(1+\gamma/s^\alpha)^\nu u^2-1, \tag{58}$$

with initial condition $u(s=0)=1/Z$. We observe that $u(s_f)=0$ if X is annihilated (i.e. $e_X^-(s_f)=0$ but $e_Y^-(s_f)>0$, and $u(s_f)=+\infty$ if Y is annihilated. For $Z=Z^*=Z^*(\gamma, \mu, \nu)$ such that (34) holds, we have a "draw," and $u^*(s)=u(s; Z=Z^*)>0$ and finite for all finite $s\geq 0$. Since (58) has a singularity at $s=0$, we apply the transformation $\tau=(s^\alpha+\gamma)^{\mu+1}$ to obtain for $\tau\geq\tau_0=\gamma^{\mu+1}$

$$du/d\tau=\tau^\beta u^2-(1-\gamma/\tau^\alpha)^\mu, \tag{59}$$

with initial condition $u(\tau=\gamma^{\mu+1})=1/Z$.

We now develop a lower bound for $Z^*(\gamma>0, \mu, \nu)$ by comparing results for $\gamma>0$ with those for $\gamma=0$. For $\gamma=0$, we denote u as w and obtain for $\tau\geq 0$

$$dw/d\tau=\tau^\beta w^2-1. \tag{60}$$

Corresponding to $Z=Z^*=Z^*(\gamma=0, \mu, \nu)$, we have via (49) and (50)

$$w^*(\tau)=\tau^{-\beta/2}K_p(T)/K_q(T)>0, \tag{61}$$

where $T=2p\tau^{(\beta+2)/2}$. Since $K_\nu(x)$ is finite and >0 for all ν, $x>0$ (see Lebedev [14], p. 136), we readily verify that $w^*(\tau)$ is finite for all $\tau>0$ and that $w^*(\tau=0)=p^{q-p}\Gamma(p)/\Gamma(q)$. Let us observe that $w(\tau_1)>w^*(\tau_1)=w(\tau_1; Z^*)\rightarrow$ we have $w(\tau)>w^*(\tau)\ \forall\ \tau\geq 0$ and $w(\tau_2)=+\infty$ for some finite $\tau_2>\tau_1$, since $D=w-w^*$ satisfies $dD/d\tau=\tau^\beta(w+w^*)D$. Consequently, $w(\tau; Z)$ corresponds to $Z<Z^*$ and $w(\tau; Z)$ becomes infinite at some finite time (see equations (49) and (50)). We now state and prove the key lemma for developing a lower bound for Z^*

LEMMA 3: Let $w^*(\tau)$ be given by (61) and let $u(\tau)=u(\tau; Z)$ satisfy (59) for $\tau\geq\tau_0=\gamma^{\mu+1}$. Then if $u(\tau_1)\geq w^*(\tau_1)$, it follows that $u(\tau)>0\ \forall\ \tau\geq\tau_1$ and $u(\tau_2)=+\infty$ for some finite $\tau_2>\tau_1$. Consequently, $Z^*(\gamma>0, \mu, \nu)>Z$.

PROOF: Consider $\tilde{D}=u-w$. It satisfies for $\tau\geq\tau_0$ via (59) and (60) the equation

$$d\tilde{D}/d\tau=\tau^\beta(u+w)\tilde{D}+\{1-(1-\gamma/\tau^\alpha)^\mu\}.$$

If $\tilde{D}(\tau_1)=u(\tau_1)-w(\tau_1)\geq 0$, then $d\tilde{D}/d\tau(\tau_1)>0$ and $\tilde{D}(\tau)>0\ \forall\ \tau>\tau_1$, Thus, when $u(\tau_1)\geq w^*(\tau_1)$, we can find $w(\tau_2)$ for $\tau_2\geq\tau_1$ such that $u(\tau_2)>w(\tau_2)>w^*(\tau_2)$, whence follows the lemma from $u(\tau)>w(\tau; \hat{Z})\ \forall\ \tau\geq\tau_2$ with $\hat{Z}<Z^*(\gamma=0, \mu, \nu)$ and the above observation that $w(\tau; \hat{Z})$ becomes infinite.

Q.E.D.

Letting $\tau_1=\tau_0$, we obtain

THEOREM 7: $Z^*(\gamma>0, \mu, \nu)>1/w^*(\gamma^{\mu+1})=\tau_0^{\beta/2}K_q(T_0)/K_p(T_0)$, where $\tau_0=\gamma^{\mu+1}$ and $T_0=2p\tau_0^{(\beta+2)/2}$. Since $w^*(\tau=0)\geq w^*(\tau)$ for $\beta\geq 0$, we have as an immediate corollary

COROLLARY 7.1: For $\nu\geq\mu$, $Z^*(\gamma>0, \mu, \nu)>p^{p-q}\Gamma(q)/\Gamma(p)$. Let us observe that for $\beta>0$, the lower bound given in Corollary 7.1 is weaker than that in Theorem 7.

9. FUTURE COMPUTATIONAL WORK

As we have seen above in Section 4, force-annihilation prediction depends on knowing the parity-condition parameter Q^*, which may be called "the Y equivalent of an X force of unit strength." We have explicitly determined Q^* (via determining Z^* for the modified auxiliary parity-condition problem (46)) for the power attrition-rate coefficients (12) in the case of no offset, i.e. $A=0$. Tabulations, for example, of the new modified exponential-like general Lanchester functions $e_X^-(s; Z^*)$ and $e_Y^-(s; Z^*)$ would facilitate force-annihilation prediction (see Theorem 5). It remains to determine Q^* for cases of positive offset, i.e. $A>0$. As discussed in Section 7, analytic results for Z^* in the modified auxiliary parity-condition problem (46) with $\gamma=A\cdot(\sqrt{k_a k_b}/(\mu+1))^{2/(\mu+\nu+2)}>0$ are apparently not possible by the usual analytical methods, so we must turn to numerical methods. It appears that a large number of cases of tactical interest (see Taylor and Brown [25]) would be covered by determining Z^* for μ, $\nu=0$, 1, 2, 3 and for a range of values of $\gamma>0$. One would be interested in, for example, plotting Z^* versus γ for fixed values of μ and ν.

Since we have developed upper and lower bounds for Z^* when $\gamma>0$ (see Theorems 6 and 7), we can use standard one-dimensional search techniques (see, for example, Wilde [28]) to calculate an approximate value of Z^* with any predetermined degree of accuracy, depending of course, on how much computation we wish to do. Since (34) must hold for all $s\geq 0$, we must determine how long (i.e. for how large a value of s) to carry out computations of $e_X^-(s; \hat{Z})$ and $e_Y^-(s; \hat{Z})$ in the modified auxiliary parity-condition problem (46) for a given trial value of Z^* (denoted as $\hat{Z}$) to see whether it is too large or too small. In the future we will show that by considering the Ricatti equation (59) one can "cut off" computations for a given value of $\hat{Z}$ well before either of two annihilation conditions (i.e. $e_X^-<0$ or $e_Y^-<0$) is actually reached.

As discussed at the end of Section 4, prediction of force annihilation within a given finite time involves the use of tabulations of the quotient of two linearly-independent general Lanchester functions. We have indicated in Section 4 that the hyperbolic-tangent-like Lanchester functions (e.g. $T_X(t)$ as defined by (42)) are to be preferred because of the accuracy of their numerical computation. Thus, there is a need for tabulations of $T_X(t)$, whose range is $[0, 1/Q^*)$ for $t\epsilon\ [0, +\infty)$. For the power attrition-rate coefficients (12) with no offset, the power Lanchester functions (also called Lanchester-Clifford-Schläfli (or LCS) functions), however, were inappropriately defined in Taylor and Brown [25] to yield such tabulations. Thus, our newer theory of force-annihilation prediction, which also involves tabulations of canonical solutions (i.e. canonical Lanchester functions) to variable-coefficient Lanchester-type equations of modern warfare, has suggested some refinements in the definition of Taylor and Brown's [25] auxiliary power Lanchester functions.† It would be desirable then to redefine the LCS functions to fit within the framework of Section 4 and to develop tabulations of their quotient. If this were to be done, linear combat models with power attrition-rate coefficients (no offset) could be analyzed with somewhat the same ease as constant-coefficient linear models (i.e. (1)).

---

†Although theoretically results for power attrition-rate coefficients with no offset are expressible in terms of "known" transcendental functions, new Lanchester functions were introduced by Taylor and Brown [25] because of lack of tabulations of these in many cases of interest.

10. EXTENSION TO A MORE GENERAL MODEL

In this section we show that all our above force-annihilation results (except those for force-annihilation within a fixed finite time) also apply to a special case of a more general model. Moreover, comparison techniques may be used to extend these results in weakened form to the general case of this more comprehensive model.

Let us consider the following Lanchester-type equations

$$(62)\qquad \begin{cases} dx/dt=-a(t)y-\beta(t)x & \text{with } x(t=0)=x_0, \\ dy/dt=-b(t)x-\alpha(t)y & \text{with } y(t=0)=y_0, \end{cases}$$

where $a(t)$, $b(t)$, $\alpha(t)$, and $\beta(t)>0$. We may think of these equations as modeling, for example, aimed-fire combat between two homogeneous infantry forces with superimposed effects of supporting weapons (which are not subject to attrition and deliver area fire against the enemy infantry). (*See* Taylor and Parry [26] for a further discussion of this model (62).) In this case, $\alpha(t)$ and $\beta(t)$ are attrition-rate coefficients which reflect the fire effectiveness of the supporting weapons [26]. Then, the force ratio $u=x/y$ satisfies the generalized Riccati equation

$$(63)\qquad du/dt=b(t)u^2+\{\alpha(t)-\beta(t)\}u-a(t) \qquad \text{with } u(t=0)=x_0/y_0.$$

For equal effectiveness of the supporting fires [i.e. $\alpha(t)=\beta(t)$], equation (63) simplifies to

$$(64)\qquad du/dt=b(t)u^2-a(t),$$

which is the same Riccati equation satisfied by the force ratio for the model (11). Hence, when $\alpha(t)=\beta(t)\forall t\geq 0$, a battle's outcome (in terms of the force ratio) is the same for the two models (11) and (62), although the battle ends more quickly for (62). Thus, in this special case all our above results on force annihilation without time limitation (e.g. Theorems 3 through 5 and their corollaries) developed for (11) (in general or with the coefficients (12)) also apply to the more general model (62). Furthermore, comparison techniques (see, for example, Hille [10]) may be used to extend these results in weakened form to (62). Consequently, we see that the force-annihilation results developed in this paper are indeed of a fundamental nature.

11. SUMMARY

We have presented a mathematical theory for predicting force annihilation for variable-coefficient Lanchester-type equations of "modern warfare" for combat between two homogeneous forces without explicitly computing force-level trajectories. † Our force-annihilation theory provides guidance for certain parameter determinations and development of tabulations of Lanchester functions (beyond those suggested in [25]) that would allow one to parametrically analyze variable-coefficient models with somewhat the same facility as constant-coefficient ones. We have shown that force annihilation can be predicted from initial conditions, without explicitly computing force-level

†In his well-known survey paper on the Lanchester theory of combat, Dolansky [9] suggested, as one of several problems for future research, developing outcome-predicting relations without solving in detail.

trajectories, by knowing a parity-condition parameter Q^*, which is the solution to a canonical auxiliary parity-condition problem. In general, this prediction would be facilitated by having tabulations of certain Lanchester functions available. The parity-condition parameter Q^* was shown to be related to the range of the quotient of two hyperbolic-like general Lanchester functions introduced by Taylor and Brown [25]. Consequently, our force-annihilation theory not only provides new information about the mathematical properties of hyperbolic-like Lanchester functions but also provides guidance for selecting canonical Lanchester functions.

We applied our general theory to the specific case of general power attrition-rate coefficients. Considering a modified auxiliary parity-condition problem, we explicitly determined Q^* (via Z^* of the modified problem) for power attrition-rate coefficients with no offset and gave upper and lower bounds for Z^* for cases of positive offset. We finally showed that certain of our force-annihilation results also applied to a more general linear differential equation combat model.

These results may be used in the analysis of the dynamic combat interactions between two homogeneous forces with time- (or range-) dependent weapon system capabilities. There is interest today in such analytic models because of improved operations research techniques for predicting Lanchester attrition-rate coefficients, in particular their temporal variations (see [2] through [5] and [25]). Further discussion of such applications may be found in Bonder and Farrell [4], Bonder and Honig [5], Taylor [22], and Taylor and Brown [25].

REFERENCES

[1] Abramowitz, M., and I. Stegun (Editors), *Handbook of Mathematical Functions*, National Bureau of Standards Applied Mathematics Series, No. 55, Washington, D.C. (1964).

[2] Bonder, S., "The Lanchester Attrition-Rate Coefficient," Operations Research, *15*, 221–232 (1967).

[3] Bonder, S., "The Mean Lanchester Attrition Rate," Operations Research, *18*, 179–181 (1970).

[4] Bonder, S., and R. Farrell (Editors), "Development of Models for Defense Systems Planning," Report No. SRL 2147 TR 70–2 (U), Systems Research Laboratory, The University of Michigan, Ann Arbor, Michigan (Sept. 1970).

[5] Bonder, S., and J. Honig, "An Analytic Model of Ground Combat: Design and Application," Proceedings U.S. Army Operations Research Symposium *10*, 319–394 (1971).

[6] Brackney, H., "The Dynamics of Military Combat," Operations Research *7*, 30–44 (1959).

[7] von Clausewitz, C., *On War*, edited with an introduction by A. Rapoport (Penguin Books, Ltd., Harmondsworth, Middlesex, England, 1968).

[8] Coddington, E., and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955).

[9] Dolansky, L., "Present State of the Lanchester Theory of Combat," Operations Research *12*, 344–358 (1964).

[10] Hille, E., *Lectures on Ordinary Differential Equations* (Addison-Wesley, Reading, Massachusetts, 1969).

[11]Ince, E., *Ordinary Differential Equations* (Longmans, Green and Co., London, 1927) (reprinted by Dover Publications, Inc., New York, 1956).

[12] Kamke, E., *Differentialgleichungen, Lösungsmethoden und Lösungen, Band 1, Gewöhnliche Differentialgleichungen, 3. Auflage* (Akademische Verlagsgesellschaft, Leipsig, 1944) (reprinted by Chelsea Publishing Co., New York, 1971).

[13] Lanchester, F. W., "Aircraft in Warfare: The Dawn of the Fourth Arm—No. V., The Principle of Concentration," Engineering, *98*, 422–423 (1914) (reprinted on pp. 2138–2148 of *The World of Mathematics, IV*, J. Newman (Editor) (Simon and Schuster, New York, 1956).

[14] Lebedev, N. N., *Special Functions and Their Applications* (Prentice-Hall, Englewood Cliffs, New Jersey, 1965) (reprinted by Dover Publications, Inc., New York, 1972).

[15] Lee, E., and L. Markus, *Foundations of Optimal Control Theory* (John Wiley, New York, 1967).

[16] McCloskey, J., "Of Horseless Carriages, Flying Machines, and Operations Research: A Tribute to Frederick William Lanchester (1868–1946)," Operations Research *4*, 141–147 (1956).

[17] Olver, F. W. J., *Asymptotics and Special Functions* (Academic Press, New York, 1974).

[18] Schwarz, H. A., "Über diejenigen Fälle, in welchen die Gaussische hypergeometrische Reihe eine algebraische Function ihres vierten Elementes darstellt," Journal für die Reine und Angewandte Mathematik (Berlin) *75*, 292–335 (1872) (also pp. 211–259 in *Gesammelte Mathematische Abhandlungen, Zweiter Band*, J. Springer, Berlin, 1890 (reprinted by Chelsea Publishing Co., New York, 1972)).

[19] Swanson, C. and V. Headley, "An Extension of Airy's Equation," SIAM Journal of Applied Mathematics, *15*, 1400–1412 (1967).

[20] Taylor, J., "A Note on the Solution to Lanchester-Type Equations with Variable Coefficients," Operations Research, *19*, 709–712 (1971).

[21] Taylor, J., "Lanchester-Type Models of Warfare and Optimal Control," Naval Research Logistics Quarterly *21*, 79–106 (1974).

[22] Taylor, J., "Solving Lanchester-Type Equations for 'Modern Warfare' with Variable Coefficients," Operations Research, *22*, 756–770 (1974).

[23] Taylor, J., "Survey on the Optimal Control of Lanchester-Type Attrition Processes," presented at the Symposium on the State-of-the-Art of Mathematics in Combat Models, June 1973 (also Tech. Report NPS55Tw74031, Naval Postgraduate School, Monterey, Calif., March 1974).

[24] Taylor, J., "Target Selection in Lanchester Combat: Heterogeneous Forces and Time-Dependent Attrition-Rate Coefficients," Naval Research Logistics Quarterly, *21*, 683–704 (1974).

[25] Taylor, J., and G. Brown, "Canonical Methods in the Solution of Variable-Coefficient Lanchester-Type Equations of Modern Warfare," Operations Research, *24*, 44–69 (1976).

[26] Taylor, J., and S. Parry, "Force-Ratio Considerations for Some Lanchester-Type Models of Warfare," Operations Research, *23*, 522–533 (1975).

[27] Weiss, H., "Lanchester-Type Models of Warfare," pp. 82–98 in *Proc. First International Conference on Operational Research* (John Wiley, New York, 1957).

[28] Wilde, D., *Optimum Seeking Methods* (Prentice-Hall, Englewood Cliffs, New Jersey, 1964).

THE DIVISIONALIZED FIRM REVISITED—A COMMENT ON ENZER'S "THE STATIC THEORY OF TRANSFER PRICING"

L. Peter Jennergren

Odense University
Odense, Denmark

ABSTRACT

Three different solutions to a very simple transfer pricing problem are outlined and contrasted. These are labeled by their authors: Hirshleifer, Enzer, and Ronen and McKinney. Weaknesses associated with each solution are pointed out.

1. THE TRANSFER PRICE PROBLEM

The topic of transfer pricing in a divisionalized firm has been discussed in a large number of journal articles and other publications over the last 20 years. One of the fundamental contributions to the transfer pricing literature is a paper by J. Hirshleifer [2], which has become the starting point for many later investigations. Hirshleifer advocated marginal cost pricing. His proposal has been largely acknowledged as a theoretically correct solution to the transfer price problem. Yet, in a recent issue of this journal, H. Enzer [1] argued emphatically that Hirshleifer's solution is incorrect. Instead, he proposed average cost pricing. This note will demonstrate that the solution proposed by Enzer is also subject to criticism. In fact, there are three different solutions, each with its own merits and faults.

The simplest possible transfer pricing situation will be considered. A company consists of two divisions, the manufacturing and the distribution division. These will be labeled Division 1 and Division 2 throughout. The company makes one product. It is manufactured in Division 1 and then transferred to Division 2 for further refinement and marketing there. There is no intermediate market. The market for the finished product is perfectly competitive, implying a fixed price level. The company wants to find a production program maximizing profit. In what follows, fixed costs will be assumed away, since they are immaterial to the analysis.

A minimum of notation must now be introduced. Let

x: amount produced in Division 1 and transferred to Division 2;

y: amount refined in Division 2 and sold to outsiders;

$C_1(x)$: total variable cost incurred in Division 1 as a function of quantity produced;

$C_2(y)$: total variable cost incurred in Division 2 as a function of quantity refined and sold to outsiders;

p: price of finished product.

The production-planning problem facing the company is then:

$$(1) \qquad \begin{cases} \text{Maximize with respect to } x \text{ and } y\text{: } py - C_1(x) - C_2(y) \\ \text{subject to: } x = y. \end{cases}$$

Assume for simplicity that $C_1(x)$ and $C_2(y)$ are differentiable, strictly convex, and increasing functions. To rule out pathological cases, it will also be assumed that:

(1) has a unique optimal solution $(\bar{x}, \bar{y}) > 0$.

Suppose it is desired to solve (1) in a decentralized fashion, meaning here that Division 1 is allowed to decide on x and Division 2 on y. But such choices cannot be made independently; and transfer pricing is one means of taking into account the interdependence of the two divisions. The relevant transfer pricing problem then becomes: How should one fix a transfer price so that it will induce Division 1 to pick $x = \bar{x}$ and Division 2 to pick $y = \bar{y}$?

2. SOLUTIONS

Three different solutions have been proposed to the transfer pricing problem: in addition to the Hirshleifer and Enzer solutions, there is also the one by J. Ronen and G. McKinney [4].

Consider a decomposition of the over-all problem (1) into divisional subproblems as follows:

(2) Division 1's problem: Maximize with respect to x: $rx - C_1(x)$;

(3) Division 2's problem: Maximize with respect to y: $py - C_2(y) - ry$.

Here, r denotes the transfer price. One can show that under the assumptions made earlier, there exists a transfer price $\bar{r}$ with the property that, for $r = \bar{r}$, $\bar{x}$ solves (2) and $\bar{y}$ solves (3) (see, for instance, Mangasarian [3], pp. 80–82); this is the basis for the Hirshleifer solution. From (2), it follows that $\bar{r}$ must be equal to $C'_1(\bar{x})$. That is, the correct transfer price is equal to marginal cost of Division 1 at the optimal production level, and that is precisely what Hirshleifer argues. To find $\bar{r}$ Hirshleifer suggests that Division 1 should inform Division 2 about its marginal cost function $C'_1(x)$. Division 2 could then determine $\bar{x}$ and $\bar{y}$ from the equation system

$$(4) \qquad \begin{cases} C'_1(x) + C'_2(y) = p, \\ x = y; \end{cases}$$

being necessary and sufficient conditions for the over-all optimal solution $(\bar{x}, \bar{y})$ to (1). $\bar{r}$ would then be determined by Division 2 as $\bar{r} = C'_1(\bar{x})$ and sent back to Division 1. With that information, Division 1 could determine its optimal production level $\bar{x}$ by solving (2) for given $r = \bar{r} = C'_1(\bar{x})$.

Against this solution, Enzer argues that Division 2 would have an opportunity to exploit Division 1. That is, the transfer price is not independent of amount acquired, x, as in problem formulation (3). Rather, the transfer price depends on x, and the relevant problem, from Division 2's point of view, is in effect the following:

$$(5) \qquad \begin{cases} \text{Maximize with respect to } x \text{ and } y\text{: } py - C_2(y) - C'_1(x)x \\ \text{subject to: } y = x. \end{cases}$$

Let the optimal solution to (5) be denoted $(\tilde{x}, \tilde{y})$, which is different from $(\bar{x}, \bar{y})$. Actually, $(\tilde{x}, \tilde{y}) < (\bar{x}, \bar{y})$, which means that Division 2 is acting as a monopsonistic buyer to exploit Division 1. Also, the transfer price becomes $\tilde{r} = C'_1(\tilde{x}) < \bar{r}$. Division 2 would send this transfer price $\tilde{r}$, rather than $\bar{r}$, to Division 1. Altogether, this results in a higher divisional profit for Division 2 but a lower divisional

profit for Division 1 and a lower total profit for the company as a whole. To avoid this situation, Division 2 may have to be outright instructed to pick $y=\bar{y}$. However, that would defeat the purpose of decentralization, according to Enzer ([1], p. 378).

To eliminate the possibility of Division 2 exploiting Division 1 while still permitting Division 2 to decide in a decentralized fashion, Enzer proposes that Division 2 should be given a price-quantity relationship other than $C'_1(x)$. Rather, Division 1 should submit to Division 2 $(C_1(x)/x)$, which is Division 1's average cost function. Taking this function as the price schedule, Division 2 constructs the divisional subproblem

$$(6) \qquad \begin{cases} \text{Maximize with respect to } x \text{ and } y\text{: } py-C_2(y)-(C_1(x)/x)x \\ \text{subject to: } y=x. \end{cases}$$

But this is obviously the same problem formulation as the over-all problem (1), meaning that decentralized decision making by Division 2 is over-all optimal. The transfer price is now $\hat{r}=(C_1(\bar{x})/\bar{x})$. This leads Enzer to conclude that average cost of the manufacturing Division 1 is the correct transfer price.

However, Enzer's solution has a drawback, too. Suppose that the transfer price $\hat{r}=(C_1(\bar{x})/\bar{x})$ is sent back to Division 1. Division 1 would then solve its problem (2) for $r=\hat{r}=(C_1(\bar{x})/\bar{x})$. The optimal solution would in general be different from $\bar{x}$. That is, if $\hat{r}$ is to be used as the one and single transfer price, then Division 1 must be instructed to pick $x=\bar{x}$. But that would also defeat the purpose of decentralization.

This suggests that two different transfer prices may be called for, one for each division. That is precisely what Ronen and McKinney advocate. The discussion up to now has been somewhat asymmetric in that Division 2 has been described as the exploiting division. However, Division 1 could equally well be thought of as exploiting Division 2. Namely, Division 1 realizes, too, that the transfer price actually paid is not independent of amount supplied. Suppose now that Division 1 is provided with the price-quantity relationship $[(py-C_2(y))/y]$, representing Division 2's demand curve for the intermediate product. Division 1 can then construct the following divisional subproblem

$$(7) \qquad \begin{cases} \text{Maximize with respect to } x \text{ and } y\text{: } [(py-C_2(y))/y]y-C_1(x) \\ \text{subject to: } x=y. \end{cases}$$

But this is obviously also the same problem as the over-all problem (1), and hence decentralized decision making by Division 1 is over-all optimal, too. The transfer price credited to Division 1 becomes $\hat{\hat{r}}=[(p\bar{y}-C_2(\bar{y}))/\bar{y}]\neq\hat{r}$. This transfer price is referred to as average revenue by Ronen and McKinney.

CONCLUSION

Three different solutions—due to Hirshleifer, Enzer, and Ronen and McKinney—to a very simple transfer pricing problem have been outlined. Each one has drawbacks:

1. Hirshleifer's solution, based on marginal cost, allows Division 2 to exploit Division 1.
2. Enzer's solution, based on average cost, eliminates this exploitation possibility, but it implies that Division 1 must be outright instructed which production level to pick.

3. The solution by Ronen and McKinney permits decentralized decision making by both divisions and eliminates exploitation possibilities. However, it is a more complex solution, since it involves two different transfer prices, based on average cost and average revenue.

One would hence have to agree with Enzer's own statement that no transfer price exists which cannot be faulted in some way ([1], p. 378). This statement applies to Enzer's own solution as well, and this author would hesitate to label the Hirshleifer solution as less correct than the other two.

REFERENCES

[1] Enzer, H., "The Static Theory of Transfer Pricing," Naval Research Logistics Quarterly, *22*, 375–389 (1975).

[2] Hirshleifer, J., "On the Economics of Transfer Pricing," Journal of Business, *29*, 172–184 (1956).

[3] Mangasarian, O., *Nonlinear Programming* (McGraw-Hill, New York, 1969).

[4] Ronen, J., and G. McKinney, "Transfer Pricing for Divisional Autonomy," Journal of Accounting Research, *8*, 99–112 (1970).

* * * * *

ADDED COMMENT

In a longer unpublished paper, from which the article published in the Naval Research Logistics Quarterly was extracted, it is suggested that the central office establish $(C_1(x)/x)$ rather than $(C_1(\bar{x})/\bar{x})$ as the transfer price. To avoid the situation Jennergren points out the recommended solution in the longer article is to treat the manufacturing division as a cost center with the objective of cost minimization. It would seem that the appropriate objective for a manufacturing division which does not sell its output in the market should be cost minimization rather than "profit" maximization. Of course, a cost center creates its own problems, such as the design of suitable incentives.

HERMANN ENZER.

A NOTE ON THE STRATEGY OF RESTRICTION AND DEGENERATE LINEAR PROGRAMMING PROBLEMS

Charles A. Holloway

Stanford University
Palo Alto, California

ABSTRACT

Using the general computational strategy of restriction, necessary conditions for optimality provide an alternative criterion for entering variables when degeneracies arises in linear programming problems. Although cycling may still occur, it is shown that if it is possible to make progress at the next iteration, the criterion is guaranteed to identify a non-basic variable which increases the value of the basic solution, thereby reducing stalling. An alternative method for determining variables to exit the basis when degeneracies occur is also suggested.

Consider the following linear programming problem:

$$\underset{x}{\text{Maximize}}\ Cx;\ \text{subject to}\ Ax=b,\quad x\geq 0 \tag{1}$$

where A is an mxn matrix of rank m. If we denote a set of basic variables by x_B, non-basic variables by x_{NB}, then the reduced costs used in the simplex criterion are given by $C=C_BA_B^{-1}A_{NB}-C_{NB}$.

Using the general computational strategy of restriction (see [3]) in which the restricted variables correspond to the non-basic variables, (1) can be written:

$$\begin{array}{lll} \underset{x}{\text{Maximize}}\ Cx;\ \text{subject to}\ Ax=b & & \underline{\text{Multipliers}} \\ x_j\geq 0 & j\epsilon F & \mu \\ x_j=0 & j\notin F & \lambda \end{array} \tag{2}$$

The computational strategy of restriction involves solving a sequence of problems in which some of the variables are set equal to zero. Each solution is tested for optimality in the original problem. If it is not optimal, then one or more restricted variables associated with negative multipliers ($\lambda_j\leq 0$), are released (added to x_j, $j\epsilon F$). If the solution to the last restricted problem resulted in an increased optimal value, any nonrestricted variable whose optimal value is zero can be added to the restricted set (see [5] for a detailed discussion of the use of restriction in concave programming).

If $\hat{x}$ is an optimal solution to (2) and μ is an optimal multiplier vector for (2), then it follows directly from the Kuhn-Tucker conditions that $\hat{x}$ is optimal in (1) if $\bar{\lambda}=\text{Minimum}\ (\mu^TA_j-C_j)\geq 0$ where A_j is the j^{th} column of A.

If we require that x_j, $j\epsilon F$ in (2) correspond to basic variables in (1), then these conditions are seen to encompass the ordinary simplex optimality conditions and $\mu=C_BA_B^{-1}$ is an optimal multiplier vector. If a solution to (2) is non-degenerate, the well known result that the simplex multipliers are unique is also easily recovered. (It can be readily verified that if non-degenerate solutions exist at each iteration, the procedure based on restriction is identical to the simplex procedure.)

When degeneracies exist, optimal multipliers are no longer unique. Under these conditions Proposition 1 yields a pricing problem which gives the directional derivative of the optimal value of (2) with respect to an increase in a restricted variable.

PROPOSITION 1: If $\hat{x}$ is an optimal solution to (2), the directional derivative of the optimal value of (2) with respect to an increase in a restricted variable is given by:

$$\Omega_F(x_j)=\underset{\mu}{\text{Minimum}}\ (C_j-\mu^TA_j)$$

$$\text{subject to } \mu^TA_j-C_j\geq 0 \qquad j\epsilon F \tag{3}$$

$$(\mu^TA_j-C_j)\hat{x}_j=0 \qquad j\epsilon F$$

PROOF: Since the optimal value of (2) is a concave function of the right-hand-side of the constraints (see, e.g., [3]), the minimum subderivative of the optimal value of (2) with respect to the right-hand-side of an equality constraint for a restricted variable is the directional derivative associated with increasing the restricted variable. Subderivatives for restricted variables are given by the corresponding components of the negative of optimal multiplier vectors, $\lambda=\mu^TA-C$ (see e.g., [4]). Therefore we have:

$$\Omega_F(x_j)=\underset{\mu}{\text{Minimum}}\ (C_j-\mu^TA_j)$$

$$\text{s.t. } \lambda_j\geq 0 \qquad j\epsilon F \tag{4}$$

$$\lambda_j=\mu^TA_j-C_j \qquad \forall j$$

$$\lambda^T\hat{x}=0$$

Expression (3) follows directly by recognizing that $\hat{x}_j=0$ $j\notin F$.||

If we define $P=[x_j|\Omega_F(x_j)>0]$, then under degenerate conditions we could choose entering variables from the set $x_j\epsilon P$ whenever $P\neq\emptyset$. As shown in Proposition 2, $P\neq\emptyset$ is a necessary and sufficient condition for an increase in the optimal value of (2) at the next iteration.

PROPOSITION 2: If the optimal solution to $(2)^t$ (superscripts denote iterations), $\hat{x}^t$, is not optimal in (1), then the solution to $(2)^{t+1}$ is such that $C\hat{x}^{t+1}>C\hat{x}^t$ if and only if $x_k\epsilon P^t\neq\emptyset$ is released from the restricted set.

PROOF: Let $x_k\epsilon P^t\neq\emptyset$ be released in $(2)^t$ forming the new restricted problem denoted $(R)^t$ in which $x_j\geq 0$, $j\epsilon F^t\cup k$ and $x_j=0$, $j\notin F^t\cup k$. If x^* is an optimal solution to $(R)^t$, then $Cx^*\geq C\hat{x}^t$ since $\hat{x}^t$ is feasible in R^t.

Assume $Cx^*=C\hat{x}^t$, making $\hat{x}^t$ an optimal solution to $(R)^t$. Then, from the Kuhn-Tucker conditions there must exist a $\bar{\mu}$ such that $\bar{\mu}^TA_j-C_j\geq 0$, $(\bar{\mu}^TA_j-C_j)\hat{x}_j{}^t=0$ for $j\epsilon F^t\cup k$. Therefore, $\bar{\mu}$ is feasible in (5) and consequently $\Omega_Ft(x_k)\leq 0$.

$$\Omega_F t(x_k) = \underset{\mu}{\text{Minimum}}\ (C_k - \mu^T A_k)$$

(5)
$$\text{s.t. } \mu^T A_j - C_j \geq 0 \qquad j \epsilon F^t$$
$$(\mu^T A_j - C_j)\hat{x}_j = 0 \qquad j \epsilon F^t$$

But $\Omega_F t(x_k) > 0$ by hypothesis and therefore $Cx^* > C\hat{x}^t$. Since $(2)^{t+1}$ is either identical to $(R)^t$ or formed by restricting variables which are at the zero level in x^*, x^* is an optimal solution to $(2)^{t+1}$ and $C\hat{x}^{t+1} > C\hat{x}^t$.

If $C\hat{x}^{t+1} > C\hat{x}^t$, then $Cx^* > C\hat{x}^t$ and by definition the variable released, x_k, must satisfy $\Omega_F t(x_k) > 0$ and hence $x_k \epsilon P^t \neq 0$.||

It follows that $P = \emptyset$ is a necessary condition for optimality (but *not* sufficient since the vector of directional derivatives is not in general a subgradient).

The strategy of restriction also provides a mechanism for determining which currently unrestricted variables should be included in the restricted set when degeneracies exist (i.e., deciding which variable should leave the basis). If at iteration t, $P^t = \emptyset$, Proposition 2 guarantees that the optimal value of the restricted problem remains the same after a restricted variable is released and therefore no variable is transferred into the restricted set. Since $\hat{x}^t$ is feasible in $(2)^{t+1}$ and $C\hat{x}^{t+1} = C\hat{x}^t$, $\hat{x}^t$ is an optimal solution to $(2)^{t+1}$ and consequently a solution to $(2)^{t+1}$ is already at hand when $P^t = \emptyset$. If $\hat{x}^t$ is not optimal in (1), then, by sequentially releasing restricted variables to $x_j, j \epsilon F$ without transferring any variables to the restricted set, we will identify an x_k such that $\Omega_F(x_k) > 0$. When such an x_k is found, the value of the objective function will increase and all variables which are in the optimal solution at the zero level can be restricted.

The purpose of this note is to illustrate the application of restriction under conditions of degeneracy. No computational advantage (in production linear programming codes) is claimed. The following example demonstrates how stalling and cycling are prevented in a special case.

An Example

This example was constructed by Beale [1] and demonstrated both cycling and stalling when the ordinary simplex procedure is used without anti-cycling devices.

(6)
$$\underset{x}{\text{Maximize}} \qquad .75x_1 - 20x_2 + .5x_3 - 6x_4$$

$$\text{subject to} \quad x_5 \quad +.25x_1 - 8x_2 - x_3 + 9x_4 = 0$$
$$x_6 \quad +.50x_1 - 12x_2 - .5x_3 + 3x_4 = 0$$
$$x_7 \quad +x_3 = 1$$

If the ordinary simplex procedure is used, x_1 is chosen for insertion into the degenerate basis (x_5, x_6, x_7) and a tie between x_5 and x_6 occurs for the variable to be dropped. If an arbitrary rule of dropping the variable with the lowest subscript is used, the seventh solution will be identical with the initial solution displayed in (6). If Charnes' perturbation scheme [2] is used to resolve ties, x_6 is

chosen and, although no progress is made this iteration, at the next iteration there is no tie for a vector to be removed, and the optimal solution is obtained in one additional iteration.

If directional derivatives are used, we solve (3) by checking $\bar{a}_{1j}$ and $\bar{a}_{2j}$ for $j=1$ and 3 ($\bar{C}_1, \bar{C}_3<0$). Since $\bar{a}_{11}, \bar{a}_{21}>0$, $\Omega_B(x_1)=-\infty$; however, $\bar{a}_{13}, \bar{a}_{23}<0$ and $\Omega_B(x_3)=.5$. Therefore, we choose x_3 to enter the basis and x_7 leaves. The revised problem becomes:

$$
\begin{aligned}
(7)\qquad & \underset{x}{\text{Maximize}} && .5 && -.5x_7 && +.75x_1 && -20x_2 && -6x_4 \\
& \text{subject to} && x_5 && +x_7 && +.25x_1 && -8x_2 && +9x_4 = 1 \\
& && x_6 && +.5x_7 && +.50x_1 && -12x_2 && +3x_4 = .5 \\
& && && x_7 && && +x_3 && = 1
\end{aligned}
$$

From (7) we see that all degeneracies have been resolved. We bring in x_1 and x_6 exits resulting in:

$$
\begin{aligned}
(8)\qquad & \underset{x}{\text{Maximize}} && 1.25 && -1.5x_6 && -1.25x_7 && -2x_2 && -10.5x_4 \\
& \text{subject to} && x_5 && -.5x_6 && +.75x_7 && -2x_2 && +7.5x_4 = .75 \\
& && && .2x_6 && +x_7 && +x_1 && -24x_2 && +6x_4 = 1 \\
& && && && x_7 && && +x_3 && = 1
\end{aligned}
$$

which is optimal.

REFERENCES

[1] Beale, E. M. L., "Cycling in the Dual Simplex Algorithm," Naval Research Logistics Quarterly *2* (4) (1955).

[2] Charnes, A., "Optimality and Degeneracy in Linear Programming," Econometrica, *20* (2) (1952).

[3] Geoffrion, A. M., "Elements of Large-Scale Mathematical Programming," I & II, Management Science, *16* (11) (1970).

[4] Geoffrion, A. M., "Duality in Nonlinear Programming: A Simplified Applications-Oriented Development," SIAM Review, *13* (1) (1971).

[5] Holloway, C. A., "A Generalized Approach to Dantzig-Wolfe Decomposition for Concave Programs," Operations Research, *21* (1) (January-February, 1973).

A NOTE ON DISTRIBUTION-FREE TOLERANCE LIMITS*

Z. Govindarajulu

University of Kentucky
Lexington, Kentucky

ABSTRACT

This note presents methods for solving for any one of the four parameters (such as sample size) involved in the construction of distribution-free tolerance limits in terms of the other three. These solutions are based on a normal approximation to the incomplete beta function. Numerical examples indicate that the approximations are very reasonable. Also considered are tolerance limits with a specified precision.

1. NOTATION, INTRODUCTION AND THE BASIC RESULT

Let $X_{1N} \leq \ldots \leq X_{NN}$ denote the order statistics in a random sample of size N drawn a continuous population having an unknown distribution function (d.f.) $F(x)$. Then it is well-known that the proportion of the population covered by the tolerance interval $(X_{r,N}, X_{N-s+1,N})$ exceeds $\beta(0<\beta<1)$ with confidence γ provided, $1-I_\beta(N-k+1, k) \geq \gamma$ where $k=r+s<N+1$, and I_β denotes Karl Pearson's incomplete beta function (See Wilks [3], p. 334). Sommerville [2] has tabulated the values of k for given N, β, and γ, and the values of γ for specified N, k and β. It is of much interest to solve for any one of N, k, β and γ when the others are specified. However, it is inconvenient to work with the incomplete beta function. In the following, we shall give a simple approximation to the incomplete beta function, which is based on the normal distribution.

PROPOSITION 1: For N sufficiently large, we have

$$I_\beta(N-k+1, k)=\Phi\left\{\frac{k-1/2-N(1-\beta)}{\{N\beta(1-\beta)\}^{1/2}}\right\} \tag{1}$$

where Φ denotes the standard normal d.f. and $I_x(a, b)$ denotes the incomplete beta function having parameters a and b.

PROOF: Let $W(N, 1-\beta)$ be a binomial random variable having parameters N and $1-\beta$. The proposition follows from the well known relationship $P(W(N, 1-\beta) \leq k-1)=I_\beta(N-k+1, k)$ and the use of the normal approximation to the binomial with the correction for continuity.

SOLUTION FOR γ: Equation (1) is already in a form to obtain γ in terms of N, k and β.

SOLUTION FOR β WHEN N, k AND γ ARE SPECIFIED: Let $z=\Phi^{-1}(1-\gamma)$, which will be negative when $\gamma>1/2$. Squaring the inequality

$$k-1/2-N(1-\beta) \leq z\sqrt{N\beta(1-\beta)}$$

*Research supported in part by the Navy under Office of Naval Research Contract No. N00014-75-C-1003 task order NR042-295. Reproduction in whole or part is permitted for any purpose of the United States Government

and solving the resultant quadratic equation leads to the solution

$$\beta=\frac{(1-a)+\frac{z^2}{2N}\pm\frac{z}{\sqrt{N}}\left\{\frac{z^2}{4N}+(1-a)a\right\}^{1/2}}{(1+Z^2/N)} \tag{2}$$

where $a=(k-1/2)/N$.

It should be noted that we should take the smaller of the two roots for our purpose since the larger root may exceed unity and we want the coefficient to be at least γ. Also when N is large,

$$\beta\doteq(1-a)+\frac{z}{\sqrt{N}}\{a(1-a)\}^{1/2}. \tag{3}$$

EXAMPLE 1: Let $k=5$, $N=100$ and $\gamma=.95$. Then $z=-1.645$ and equation (2) yields $\beta=.907$ and equation (3) gives $\beta=.921$, whereas the true value of β is .90 (See Sommerville [2]).

SOLUTION FOR N: Let $N(1-\beta)=M$ and let $z=\Phi^{-1}(1-\gamma)$, and solving for $\sqrt{M}$ from the inequality

$$k-1/2-M\leq z\sqrt{\beta M}.$$

we obtain

$$\sqrt{M}=[-z\sqrt{\beta}\pm\{\beta z^2+4(k-1/2)\}^{1/2}]/2.$$

Thus

$$\mathrm{N}=[-z\sqrt{\beta}+\{\beta z^2+4(k-1/2)\}^{1/2}]^2/4(1-\beta). \tag{4}$$

EXAMPLE 2: Set $\gamma=.95$, $\beta=.9$, $k=9$, $z=-1.645$.

Then

$$\sqrt{M}=(1.5604\pm6.036)/2.$$

Here we take the larger of the two roots (so as to be conservative) and get

$$\sqrt{M}=3.798, \text{ yielding } N=144.25$$

whereas the true value from Sommerville [2] is 144.

SOLUTION FOR k: Let $z=\Phi^{-1}(1-\gamma)$.

Then

$$k\leq z\{N\beta(1-\beta)\}^{1/2}+N(1-\beta)+1/2. \tag{5}$$

In the following we tabulate the value of k for specified values of N, β, and γ and compare them with the true values obtained by Sommerville [2]. The numbers in parentheses represent the true values given by Sommerville [2].

N	$\gamma=.99,\ \beta=.95,\ k\leq .5+(.05)N-(.51)N^{1/2}$	$\gamma=.95,\ \beta=.90,\ k\leq .5+(.1)N-(.49)N^{1/2}$
100	.42 (1)	5.57 (5)
121	.96 (1)	7.17 (7)
144	1.61 (2)	8.98 (9)
169	2.35 (3)	10.98 (11)
400	10.35 (11)	30.63 (30)
625	19.06 (19)	50.65 (50.5)

2. "GENERALIZED" TOLERANCE LIMITS

One can relate the tolerance and confidence limits by considering the generalized tolerance limits proposed in the following Let $U_{r,s,N}=F(X_{N-s+1,N})-F(X_{r,N})$. Also let $\alpha=EU_{r,s,N}=1-k/(N+1)$ where $k=r+s$. Consider the following relation:

(i) $$P(|U_{r,s,N}-\alpha|\leq\Delta)\geq\gamma,$$

(ii) $$P(U_{r,s,N}-\alpha\geq-\Delta)\geq\gamma,$$

or

(iii) $$P(U_{r,s,N}-\alpha\leq\Delta)\geq\gamma.$$

Then one should be able to solve for one of the quantities α, Δ, γ and N in terms of the rest. This formulation is somewhat appealing since it is analogous to setting up one-sided or two-sided fixed-width confidence intervals for α with specified γ. However, α does not act like an unknown parameter since its value will be known as soon as $r+s$ and N are specified. Using methods analogous to those of Section 1, one can solve for Δ for specified α, γ and N. For further details the reader is referred to the author's (1976) technical report [1].

3. TOLERANCE LIMITS HAVING A SPECIFIED PRECISION

Recall that the proportion of the population covered by $(X_{r,N},\ X_{N-s+1,N})$ is

$$U_{r,s,N}=F(X_{N-s+1,N})-F(X_{r,N})=U_{N-s+1,N}-U_{r,N}$$

where $U_{1N}\leq\ \ldots\ \leq U_{NN}$ denotes the order statistics in a random sample of size N drawn from the uniform (0, 1)-population. Straightforward computations yield

(6) $$EU_{r,s,N}=1-k(N+1)^{-1},\qquad k=r+s$$

(7) $$\text{Var}\ U_{r,s,N}=k(N+1-k)/(N+1)^2(N+2)\doteq\alpha(1-\alpha)/(N+2)$$

where $1-\alpha=EU_{r,s,N}$.

Letting Var $U_{r,s,N} \leq b$ is equivalent to either

$$\delta = 1-\alpha \geq (1/2)[1+\{1-4b(N+2)\}^{1/2}] \tag{8}$$

or

$$\delta \leq (1/2)[1-\{1-4b(N+2)\}^{1/2}]$$

where we assume that $b \leq 1/4(N+2)$.

EXAMPLE 3: Let $b=.002$, $N=99$; this implies that either $k \geq 72$ or $k \leq 28$.

When γ, β and γ are specified one can solve for N in the following manner:

$$P(U_{r,s,N} \geq \beta) \geq \gamma)$$

implies after using equation (1) that

$$k \leq z\{N\beta(1-\beta)\}^{1/2}+N(1-\beta)+1/2, \qquad z=\Phi^{-1}(1-\gamma).$$

That is,

$$1-\alpha \doteq \frac{k}{N} \leq \frac{z}{\sqrt{N}}[\beta(1-\beta)]^{1/2}+(1-\beta)+\frac{1}{2N}.$$

Ignoring $(2N)^{-1}$ and solving for N we have

$$N^{1/2} \geq -z\{\beta(1-\beta)\}^{1/2}/(\alpha-\beta), \qquad \alpha > \beta,$$

or

$$N \geq z^2[\beta(1-\beta)]/(\alpha-\beta)^2. \tag{9}$$

EXAMPLE 4: Let $\alpha=.95$, $\beta=.9$, $\gamma=.95$. Then $z=-1.645$ and equation (17) gives $N \geq 98$.

ACKNOWLEDGMENT

I thank the referee for a critical reading of the paper.

REFERENCES

[1] Govindarajulu, Z., *A Note on Distribution-free Tolerance Limits*, University of Kentucky Department of Statistics, Technical Report No. 97 (1976).

[2] Sommerville, P.N. Tables for obtaining nonparametric tolerance limits, Annals of Mathematica Statistics, 29, 599–601 (1958).

[3] Wilks, S.S., Mathematical Statistics (John Wiley and Sons, New York, 1962).

U.S. GOVERNMENT PRINTING OFFICE: 1977 240-830/2 1-

INFORMATION FOR CONTRIBUTORS

The NAVAL RESEARCH LOGISTICS QUARTERLY is devoted to the dissemination of scientific information in logistics and will publish research and expository papers, including those in certain areas of mathematics, statistics, and economics, relevant to the over-all effort to improve the efficiency and effectiveness of logistics operations.

Manuscripts and other items for publication should be sent to The Managing Editor, NAVAL RESEARCH LOGISTICS QUARTERLY, Office of Naval Research, Arlington, Va. 22217. Each manuscript which is considered to be suitable material for the QUARTERLY is sent to one or more referees.

Manuscripts submitted for publication should be typewritten, double-spaced, and the author should retain a copy. Refereeing may be expedited if an extra copy of the manuscript is submitted with the original.

A short abstract (not over 400 words) should accompany each manuscript. This will appear at the head of the published paper in the QUARTERLY.

There is no authorization for compensation to authors for papers which have been accepted for publication. Authors will receive 250 reprints of their published papers.

Readers are invited to submit to the Managing Editor items of general interest in the field of logistics, for possible publication in the NEWS AND MEMORANDA or NOTES sections of the QUARTERLY.

NAVAL RESEARCH
LOGISTICS
QUARTERLY

JUNE 1977
VOL. 24, NO. 2
NAVSO P-1278

CONTENTS

ARTICLES		Page
À Priori Error Bounds for Procurement Commodity Aggregation in Logistics Planning Models	A. M. GEOFFRION	201
A Node Covering Algorithm	E. BALAS H. SAMUELSSON	213
Two-Characteristic Markov-Type Manpower Flow Models	W. J. HAYNE K. T. MARSHALL	235
Joint Pricing and Ordering Policy for Exponentially Decaying Inventory with Known Demand	M. A. COHEN	257
Estimation of Ordered Parameters from k Stochastically Increasing Distributions	H. J. CHEN	269
An M/M/1 Queue with Delayed Feedback	I. HANNIBALSSON R. L. DISNEY	281
Optimal Dynamic Rules for Assigning Customers to Servers in a Heterogeneous Queuing System	W. WINSTON	293
Computing Bounds for the Optimal Value in Linear Programming	M. KALLIO	301
Solving Multicommodity Transportation Problems Using a Primal Partitioning Simplex Technique	J. L. KENNINGTON	30
Stochastic Transportation Problems and Other Network Related Convex Problems	L. COOPER L. J. LEBLANC	32
Probabilistic Ideas in Heuristic Algorithms for Solving Some Scheduling Problems	I. GERTSBACH	33
Force-Annihilation Conditions for Variable-Coefficient Lanchester-Type Equations of Modern Warfare	J. G. TAYLOR C. COMSTOCK	34
The Divisionalized Firm Revisited—A Comment on Enzer's "The Static Theory of Transfer Pricing"	L. P. JENNERGREN	37
A Note on the Strategy of Restriction and Degenerate Linear Programming Problems	C. A. HOLLOWAY	37
A Note on Distribution-Free Tolerance Limits	Z. GOVINDARAJULU	3

OFFICE OF NAVAL RESEARCH
Arlington, Va. 22217